bls Berkeley Linguistics Society

Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Legacy of Grice (1990)

The Annual Proceedings of the Berkeley Linguistics Society is published online via eLanguage, the Linguistic Society of America's digital publishing platform.

PROCEEDINGS OF THE SIXTEENTH ANNUAL MEETING

OF THE

BERKELEY LINGUISTICS SOCIETY

February 16-19, 1990

GENERAL SESSION

and

PARASESSION

ON

THE LEGACY

OF

GRICE

Berkeley Linguistics Society
Berkeley, California, USA

PROCEEDINGS OF THE SIXTEENTH ANNUAL MEETING

OF THE

BERKELEY LINGUISTICS SOCIETY

February 16-19, 1990

GENERAL SESSION

and

PARASESSION

ON

THE LEGACY

OF

GRICE

edited by

Kira Hall

Jean-Pierre Koenig

Michael Meacham

Sondra Reinman

Laurel A. Sutton

We respectfully dedicate this volume to the memory of H. P. Grice.

TABLE OF CONTENTS

GENERAL SESSION

vii

PARASESSION

Bulging monosyllables: Areal tendencies in Southeast Asian diachrony   543
  JAMES A. MATISOFF

<div align="center">•   •   •</div>

The following papers were presented at the conference but do not appear in this volume:

The girl blue her eyes
  ORIN GENSLER

The cognitive consequences of linguistic change
  WILLIAM LABOV

The following papers were not presented at the conference but are included in this volume:

Quantifier float in Korean
  KI-SUN HONG

Bulging monosyllables: Areal tendencies in Southeast Asian diachrony
  JAMES A. MATISOFF

From tone to stress: Mechanisms and motivations
  JOE SALMONS

ix

SPECIAL SESSION

GENERAL SESSION

Rethinking conversational code-switching:
codes, speech varieties, and contextualization

Celso Alvarez-Cáccamo
University of California, Berkeley

1. INTRODUCTION [1]

The issues that will concern me in this paper are the treatment of code-switching as a signalling device in face-to-face interaction; the identification which is commonly established between SPEECH VARIETIES and CODES; and the conceptualization of codes as supposedly discrete sets of 'co-occurrent linguistic features' (Ervin-Tripp 1973) or 'clusters of co-occurrent variables' (Gumperz and Herasimchuk 1975).

From Weinreich's (1953) pioneering reference to the phenomenon of 'switching codes' (:73), to Gumperz's instrumental work on the metaphorical attributes of code choices (1964, 1970, 1982; Blom and Gumperz 1972; Gumperz and Hernandez-Chavez 1971), to Gal's (1987) [2] materialist treatment of language choices as a reflection of the group's class-structural position, the extensive sociolinguistic literature on code-switching has revealed that much social, interactional, discursive, and other information can be signalled through the alternation of speech varieties in their broadest sense.

In general terms, code-switching introduces socio-cultural information in context, which is retrievable through conversational inference (Gumperz 1982). At the level of social organization, code choice patterns have been argued to: reflect social structure, specifically class-structural positions (Gal 1987; Hill 1985; McClure and McClure 1988), and changing interethnic relationships (Heller 1982; Woolard 1983b); signal group membership, in particular local identification (Blom and Gumperz 1972, Gumperz 1982), ethnicity (Gumperz 1982; Mitchell-Kernan 1972), and gender (Gal 1978); be correlated with group roles of leadership and subordination (Calsamiglia and Tusón 1980); channel 'the speaker's claim to a social status' (Gal 1979:171); constitute a socio-functional 'style' (Gumperz 1964), a culturally specific 'mode of speaking' (Poplack 1980), or an 'unmarked choice' among multilinguals (Scotton 1988); manage the speaker's 'ambiguous' or dual group identification (Heller 1982, 1983; Scotton 1976, 1983, 1988); and, in general, 'invoke' (Irvine 1979) social identities in discourse, such as official personae vs. private identities (Alvarez Cáccamo 1989a, to appear; Blom and Gumperz 1972, Irvine 1979, 1982).

At the interactional and sequential levels, code choices have been put in connection with: the performance of specific speech activities (Blom and Gumperz 1972; Mitchell-Kernan 1972) and discourse tasks (Auer 1984; Gumperz 1982); the selection of an addressee (Auer 1984; Duranti, to appear; Gumperz 1982); the management of requests (Gumperz 1982, Valdés 1981), disputes (Gal 1979) and interactional conflict (Duranti, to appear); the management of narrative voices in reported speech (Alvarez Cáccamo 1989a; Mitchell-Kernan 1972) and in humor (Woolard 1983a); and the speaker's signalling of 'epistemological dispositions' (Ochs, to appear) and affect (Irvine 1982).

Finally, at the superstructural level, code choices may symbolize ideology (Mitchell-Kernan 1972), the community's 'cultural philosophy and value system' (Albert 1972), and political ideologies and positions (Alvarez Cáccamo 1989a, to appear).

However, from most of this research one gets the impression that whenever we find language or register alternation in discourse, each speech variety in contrast carries its social and discursive meanings through its unmistakable association with categories such as ethnicity, class, or gender. For instance, with regard to social identity, it has been argued that a speaker may claim two (or more) distinct roles or social identities by switching languages or dialects for specific situations (Blom and Gumperz's [1972] 'situational' switching), a dual identity through repeated intrasituational switching (Heller's [1982] 'negotiations' of identity; Poplack's [1980] 'code-switching mode'; Scotton's [1988] 'unmarked choices'), or a single identity by never switching even when being able to do so (Gal's [1979] 'unreciprocal language choices', Woolard's [1983] 'Bilingual Norm'). From this perspective, each variety is generally presented as constituting a discrete code, internally bound by co-occurrence constraints, and paradigmatically contrasted with other varieties-codes. Additionally, often phonological and, particularly, prosodic variations are left unexamined, and languages in alternation are practically treated as indivisible constructs, devoid of any internal variation.[3]

I believe that these views about the systematic correlations between linguistic varieties and social meanings need to be reexamined. My feeling is that often researchers impose upon the observed interaction their preconceived notions about what constitutes a code as a social-indexical system. However, as in many other aspects of the study of language in action, the linguist's or sociolinguist's categories may not match the participants' own definition of the situation.

In my research on bilingual conversation and language revival in Galiza, Spain, I have registered a number of salient cases of language alternation that somehow defy standard interpretations of code-switching. At times it is hard to determine what exactly is being signalled through a specific switch, although it can be seen, after detailed examination of conversational organization, that changes have taken place in terms of activity framing or participants' alignments. Other times, switches seem totally meaningless and unmotivated; they may be connected to some rhythmic properties of talk which I have not explored fully. Yet in other cases, several, simultaneous threads of signification run parallel in code choices. Phonological and prosodic registers, languages, dialects, and styles intertwine in a web of situated meanings, thus concertedly opening a space for interpretation of what is being said.

In this paper I claim that it is the co-occurrence and interplay of these specific signalling systems, and not the presence of any one in isolation, that constitute specific COMMUNICATIVE CODES which contextualize meaning. I will examine two cases where, in the constant recontextualization of the speech exchange, shifts between speech varieties cannot be mapped exactly against the transitions between activities and the frame shifts which do occur. In my analysis, I draw on Gumperz's (1982, 1989) instrumental view of code-switching as a CONTEXTUALIZATION CUE for the interpretation of messages.[4] I have specifically asked what (if anything) is contextualized in each case of alternation of speech

varieties, and, particularly, how it is contextualized. Ultimately, my aim is to contribute to a broad, dynamic conceptualization of code-switching which may give account of various sorts of alternations between speech varieties, registers, languages, variants, or single markers in discourse.

2. OPENING THE FRAME

Case 1, which I have titled 'Opening the Frame', exemplifies the subtle turns that the situated manipulation of varieties may take for the constitution of communicative codes and discursive voices. The episode comes from the meeting of the board of directors of a museum run by the city where I did my research in southwestern Galiza. The languages of the entire meeting were Galician and Spanish alternately.

The nine members of the board are about to start to discuss the last point in the agenda. The entire episode from lines 1 to 32 can be seen as an off-record, pre-procedural exchange. At the onset, A, indirectly supported by B, expresses his interest in going ahead with the discussion expediently (lines 1-6), as apparently the members of the board (all of them male) wanted to return home in time to watch a soccer game between Spain and Yugoslavia that was being broadcast on television that evening. To A's and B's suggestion, the Chairman (C) and the Secretary (S) of the board insist cooperatively that there is still time to go ahead with the discussion, as the game would not be starting until a few minutes later (8-13).

In reference to the soccer game, an activity emerges (17-32) where one of the participants, D launches a humorous frame by quoting the remarks made by a Portuguese sports commentator about the Yugoslavian team in light of the upcoming match. After an initial attempt to take the floor (14), D comments that the Portuguese journalist had referred to the Yugoslavs as *'very dangerous communists'* (19). The others laugh, presumably at the associations that the sportscaster had made between the unbeatable Yugoslavian soccer team and the Yugoslavian political regime.[5] Finally (33), the meeting resumes with the discussion of the last point of the agenda preceded by a report (in Spanish, coincidentally, but not significantly) by the speaker in charge. Spanish is represented by the Helvetica font, and Galician goes in *Times Roman Italic:*[6]

1. Opening the frame

1	A	esto va rápido {[hi] hombre} //	A	This will go fast, come on.	
2		que queremos ir al fútbol todos //		We all want to go (watch) the soccer game.	
3	?	xx	?	xx	
4	A	no es cierto que queremos ir al fútbol ?	A	Isn't it true that we all want to watch the game?	
5	B	que hoy es un buen partido / yugoesla:via /	B	(Yes,) today's is a great game. Yugoslavia.	
6		no es -		It's not ——	
7	?	si /	?	Yeah.	
8	C	*si pero empeça às *nove /*	C	*Yes, but it starts at nine.*	
9		*inda temos: - som as oito e: .. vinte ,*		*We still have —— it's eight-twenty.*	

10	S	⌐nos queda media hora⌐ /	S ⌐There is still half-hour left.
11	C	⌐inda temos um ⌐ -	C ⌐We still have a —
12		inda temos um {[ac] pouquinho de tempo } /	we still have a little time.
13		{[lo] (venha) } /	Let's start.
14	D	como di⌐ -	D As (he) sa(id) —
15	B	⌐tú lo que pasa es que hay que levan'tar⌐se ,	B Yeah, but the thing is, we have to get uup...
16	E	⌐bueno /	E Come on.
17	D	⌐como di⌐zía o::: cronista portughés ↑	D As the Portuguese (sports) commentator said,
18		os iugoslavos ,	'Those Yugoslavs?
19		{[hi] comunistas peligrosíssimos } /	Very dangerous communists!'
20	C	⌐mja ja JA JA JA JA JA	C ⌐ha ha HA HA HA HA HA
21	?	⌐ ja ja ja	? ⌐ ha ha ha
22	F	⌐ ahh	F ⌐ ahh
23	A	⌐si si si si /	A ⌐Right, right, right.
24	B	⌐xx xx portugueses /	B ⌐xx xx Portuguese (pl.).
25	D	dizia o cronista portu⌐gués ,	D (That) said the Portuguese commentator.
26	?	⌐je JE JE JE JE JE	? ha HA HA HA HA HA
27	B	comunistas peligrossísimos /	B 'Very dangerous communists'.
28	D	comunistas peligrosíssimos / hh	D 'Very dangerous communists'. hh.
29	F	hh je JE JE JE JE JE	F hh ha HA HA HA HA
30	?	.h je je .h.h je je .h.h je je	? .h ha ha .h.h ha ha .h.h ha ha
31	B	pues ganaron /	B Well, they won.
32	?	je je je	? ha ha ha
33	E	... PARtimos de un esquema ,	E WE START from a scheme
34		que el otro día ,	that the other day
35		hablamos aquí ,	we discussed here,
36		y: que lo recogí inmediatamente al día siguiente ,	and that I included immediately the next day...

Let us now focus on the linguistic resources employed by D in his interactional maneuvers. In his introduction to the quote ('as the Portuguese (sports) commentator said', 19) D uses a variety of Galician known as galego de gheada (represented by the underlined italics). This variety is a non-standard phonological register characterized by gheada [heáða], that is, the fricativization and devoicing of the voiced velar stop: /g/ -> [ɣ] (other variants are [x], [h]; for a review of the phenomenon see for example Alvarez Cáccamo 1989b:261-65). Portughés [portuhés], then, is one of the gheada pronunciations of standard portugués [portuɣés]. The devoiced variant is commonly represented by gh.

As a stigmatized socio-geographic variety, gheada Galician is the native dialect of farming and fishing villages and, in general, of the working classes all throughout southwestern Galiza. Through schooling and the exposure to standard Spanish — and, since recently, to standard Galician — educated speakers gradually relegate their native variety to intimate, in-group functions. At the same time, they construct new meanings for gheada — among them, the feeling that it is, at its best, a form to use colloquially, not for formal purposes. Nevertheless, even educated

urban dwellers who did not acquire *gheada* natively are aware of its socio-stylistic meanings for local identification and informality, and they deploy *gheada* strategically in their daily talk.[7]

To be noted in Case 1 is that the humorous frame opens with a line in *gheada* Galician, but there is no further use of such variety in the episode. Notice how after the joke is delivered, D repeats his introduction to the enacted quote '*(That) said the Portuguese commentator*' (25), but this time he uses standard Galician.

I want to emphasize that, in the context, *gheada* does not primarily function to invoke D's social background, or his class affiliation or consciousness. Nor does *gheada* simply signal, as one would be tempted to interpret the exchange, less communicative distance among participants than non-*gheada* Galician. While presumably these meanings somehow underlie the contextualized interpretation of the exchange, here *gheada* is primarily a <u>metaphorical</u> device, whose deployment functions, through intricate inferential associations, to cue into a new frame or space for interpretation of what is to come. Like speech markers (Brown and Levinson 1979), choices of speech varieties may work in non-congruent ways to signal indirectly other than their basic associative meanings. In Case 1, significantly, once the frame for a joking exchange has been opened (and thus, I argue, once the <u>code</u> has been switched) it is unimportant what variety of Galician (or perhaps Spanish) continues to be used.

Thus, while paradigmatic contrasts are established between at least three recognizable forms of speech (namely, Spanish, *gheada* Galician, and standard Galician), overall, in the episode the participants only engage in two relevantly distinct activities: (1) informal procedural talk; and (2) topically connected to the prior activity, joking talk including reported speech as a specific discourse task.

3. SPEAKING GALICIAN IN SPANISH

Case 2 exemplifies the subtle interplay of simultaneous signalling systems in talk. With the analysis of these data, I want to argue that we need to examine closely this interplay in order to identify the communicative codes that are actually being constructed for the expression of local meanings according to members' methods and perceptions.

The excerpt corresponds to a television interview with the president of the Galician Federation of Sport Shooting. The official (P) has been asked about the new facilities for practicing airgun shooting which had been built in a given sporting club with financial support from the Federation. In his answer, P begins by reviewing the importance of the facilities for the development of this sport in Galiza. Then (off the transcript) he lists the alleged virtues of airgun shooting for the youngsters as a recreational activity (e.g. 'coordination and self-control'). Finally (93-108), he goes on to complain about the status of airgun shooting in the curriculum of high-school sport activities in Galiza.

I will focus specifically on this latter activity, where at least two systems of signification run parallel in P's speech. But, first, it is relevant to examine briefly P's overall performance throughout the interview. The beginning of the official's answer (69-72) exemplifies the characteristic ambiguity he displayed in his conversational orientation toward the interviewer and the audience. In playing with social voices,

the interviewee systematically began each answer in Galician. In each case, he did so by recapturing (almost like <u>mentioning</u>) the question posed by the Galician-speaking interviewer (69). After one or two utterances, P altered his stance, to invoke his official, authoritative public persona by using Spanish (69-72):

2(a). *Speaking Galician in Spanish*

67	I	*e::h senhor #last name# /* *como último: pregunta ,*	I	*Uh, Mr. #last name#,* *one last question.*	
68		*qué supóm pró clu (=clube)* *estas novas instalaciós /*		*What does* [8] *these new facilities* *represent for the Club?*	
69	P	*pos estas novas instalaciós* *presupone -*	P	*Well, these new facilities* *presupposes ——*	
70		*pues *muy *muy importante /*		*well, very, very important,*	
71		*que el tiro neumático ,*		*since airgun shooting*	
72		*se acaba de declarar como* *decía hace un momento* *olímpico* ↑		*has just been declared, as* *I just said a minute ago,* *(an) Olympic (sport)...*	

Thus, a certain ambiguity detectably colors P's performance. Additionally, as we will see, at specific points in the interview it cannot be determined exactly what language P is speaking. In sum, one perceives that, overall, P is coming across both as the representative of an institution, and as one of us, a plain Galician guy.

The manipulation of verbal resources for this sort of sociolinguistic illusionism is most salient in the end of the episode:

2(b)

		[several turns omitted]	
93	P	.h.h y es [ə] es muy importante ↑	.h.h And it is, uh, it is very important
94		{[ac] que por ejemplo el tiro neumático } a nivel europeo ↑	the fact that for example all around Europe airgun shooting
95		esté 'incluído como deporte esco*lar ,	is included as a high school sport
96		en todos [ə] los: centros de {[hi] ense*ñanza } /	in all uh, the educational centers.
97		*e sem embargo {[hi] ei*qui } ↑*	*And yet here,*
98		*.h.h.h como somos assi tam* *{[hi] es:'pe*ciales }↑*	*.h.h.h since we are like, so* *'peculiar',*
99		*pos resulta que em {[hi] galicia } ↑*	*well, it turns out that in Galicia,*
100		eh que hasta el tiro neumático siempre estaba dentro d- del deporte {[hi] escolar } ↑	uh, where airgun shooting had always been included as a school sport,
101		ahora pos lo han - lo han {[hi] quitao } ↑	well, now they've, they've done away with it,
102		y es una verdadera pena ↑	and this is really a shame
103		.h.h porque el número de: chavalitos a partir de los once o doce años que pueden .h participar ,	.h.h because the number of boys eleven-years old or twelve-years old or older who can .h participate
104		podía ser *muy *grande ↑	could be very large,
105		y podían tener una actividad ,	and they could enjoy an activity

106	y un entretenimiento ,	and an entertainment,
107	{[hi] y un futuro } ↑	and a future
108	*muy bueno //	(all) very good.

P's report in Spanish about sport shooting in high school in Europe (93-96) must be viewed as the positive term of the opposition that he establishes immediately after, in what constitutes his climactic complaint (97-102): *'and yet here, since we are like, so "peculiar", well, it turns out that in Galicia, uuh, where airgun shooting had always been included as a school sport, well, now they've, they've done away with it, and this is really a shame'*.

Notice, firstly, that the utterance in line 96, while contained in a reportive assessment, already shows the prosodic pattern characteristic of the next turns — an extra high pitch over the last lexical item. Additionally, in the sequence from lines 96 to 101, this higher-pitch word at times carries the phrase accent (represented by the asterisk). If we examine the ordering of these emphasized items, the thematic organization of P's complaint becomes neatly summarized in a sequential nutshell around the implicit object 'Sport Shooting':

ENSEÑANZA-*EIQUI*-*ESPECIALES*-*GALICIA*-ESCOLAR-QUITAO
'education-*here*-*peculiar* (pl.)-*Galicia*-educational-taken away'.

A major shift to a local, 'we' voice takes place in line 97, *'and yet here'*. Language choice and local identity match unambiguously in dialectal southern Galician *eiqui* 'here' (standard Galician *aqui*). In the utterances that follow, however, Galician and Spanish converge to the extent that it is absolutely impossible to determine what language P is speaking. Both in *2(b)* and in *3*, I have transcribed this syncretic stretch of discourse in the *italicized Helvetica font*:

3

98	*komo sómos asI táη espeθjáles*
99	*pọs resúlta ke eη galíθja* [9]

To be sure, what language is being used is largely irrelevant in the exchange; the issue is how language is being used, in an implicitly evaluative stretch of discourse internally kept together by prosodic regularities, not by language. Similarly, thematic cohesion in lines 97-99 is established not through language choices, but through lexical references to in-group membership: *'here'* (97), *'we are'* (98), and *'Galicia'* (99). Finally, the sequence in lines 100-101, while undoubtedly Spanish, is at the same time linked to the previous turns by prosody.

Briefly, P's performance presents (a) on the one hand, a gradual transition between lexical and syntactic registers (from Spanish to Galician to an ambiguous register again to Spanish); and (b) on the other, the juxtaposition of two prosodic patterns (one with a final extra-high pitch, the other showing a flatter contour). The syntactic and iconic (cf. Bolinger 1985) lines of signification run parallel to each other, but they do not overlap exactly, as can be seen in *4*. Single underlining stands

for the flat prosodic register, and <u>double underlining</u> indicates the emphatic, evaluative prosody:

4

```
93    y es muy importante ↑
94    que por ejemplo el tiro neumático a nivel europeo ↑
95    esté incluído como deporte escolar ,
96    en todos los centros de enseñanza /
97    e sem embargo eiqui ↑
98    komo sómos así tán espeθiáles ↑
99    pos resúlta ke en galíθja ↑
100   que hasta el tiro neumático siempre estaba dentro del deporte escolar ↑
101   ahora pos lo han quitao ↑
102   y es una verdadera pena ↑
```

The sequence above graphically illustrates the <u>crescendo</u> and subsequent <u>diminuendo</u> in which P engages, in his breakthrough from an expository frame to an evaluative one, through the combination of prosody, lexis, and syntax. Notice, indeed, how P's final, explicit assessment 'this is really a shame' (102) again shows an expository prosody, as if to minimize the speaker's personal involvement in what now sounds not like a complaint, but like the statement of a regrettable reality.

Research on code-switching (particularly Heller's and Scotton's work) has highlighted the role of language choices in managing the speaker's ambiguous roles and social identifications. This work has focused on languages as indivisible constructs. But what can be observed in the previous cases is the interplay of several aspects of the speaker's repertoire to construct social and discursive voices, in Hill's (1985) sense. In the actualization of potential meanings (cf. Voloshinov 1973 [1929], Gumperz and Herasimchuk 1975), the vernacular prosody analogically invokes the speaker's personal, rather than positional identity in his indirect complaint. On its part, the expository prosody alludes in this case, in combination with Spanish, to the speaker's official persona. And all throughout the event, P's quick juxtaposition of Spanish and Galician and his deployment of ambiguous utterances constitute skillful procedures to contextualize the tone of the exchange and to establish situated tactical alliances with the interviewer and the audience.

In making sense of Case 2, I would like to report on the testimony of one of my informants who was presented with the original recording for interpretation. While all of the native speakers consulted viewed some sort of connection between the use of Galician and the claim to local identity (as in line 97), an informant offered a particularly important clue about the use of Spanish with a vernacular prosody (98-101). She referred to this sequence as '*Spanish with a Galician attitude*' (*anque é em espanhol, é co'a actitú galega*). In other words, one may say that the language of P's complaint is Spanish, but the <u>spirit</u> is Galician — it is <u>Galician with Spanish words.</u>

Thus, one can speak Galician <u>in</u> Spanish. It is not language per se, but this *actitú* or 'spirit' of a social voice that I associate in this case with a code — a symbolic frame of interpretation. In the context, each of the two codes (one

reportive, the other evaluative) contains elements of two languages; simultaneously, each code presents a distinct prosodic pattern. The reportive code is socially associated with Spanish; the evaluative/complaining code has been equated with a 'Galician attitude'. The use of each code alludes to a represented social identity, and retrieves for social interaction a given set of mutual dispositions — from in-group camaraderie to official distance. Similarly, Case 1 showed a code for informal procedural talk (contextually associated with Spanish and Galician alike) and a code for jovial conversational involvement (associated with two distinct varieties of Galician). In other contexts one may find narrative codes, authoritative codes, requesting codes, trouble-telling codes, and so on, and the associations between codes and their constitutive varieties may be of a different nature.

4. CONCLUSION

Consequently, what looks like a linguistic code for the linguist may not <u>count</u> as a communicative code for conversationalists. One of the objectives of the interpretive study of social interaction ought to be, precisely, to reveal what counts as a communicative code in a given encounter. In a provisional, rough formulation, I have thus viewed each code both as a frame for interpretation and as a system of linguistic production. As a frame for interpretation, each code is a universe of significations collectively constructed on the basis of transituational experience. As systems of linguistic production, codes are internally bound by co-occurrence constraints which govern what can and cannot be uttered within their confines. Note, however, that these constraints are situational and discursive, not linguistic or structural, so that what confers coherence to the code is the speech situation or activity itself. Codes may thus be constituted by elements that cut across seemingly distinct speech varieties, as in the cases reviewed. Consequently, the situated meanings of communicative codes derive <u>only indirectly</u> from the social meanings of the constituting linguistic varieties.

It goes without saying that, in the study of social actors' integrated communicative repertoire, the syntactic should not be separated from the prosodic, the lexical, the gestural or the kinesic on the basis of pre-conceived notions of what constitutes a code. I have stressed the need to refocus our analysis on the native categories and methods that people put at work in social encounters to construct communicative codes. From this approach, the act of code-switching is not necessarily the act of alternating between speech varieties, but the act of shifting gears in communicative behavior, the act of <u>micro-chronologically recontextualizing talk and reality</u>. This is often accomplished by switching languages, dialects, styles, or registers — and all the literature points in this direction. Further, it may also be accomplished by the use of single markers such as honorifics or phonological variants, which constitute the visible 'tip of the iceberg' of the diversity of broadly overlapping codes in social life. But, importantly, the act of switching communicative codes may also be accomplished by <u>not</u> alternating between languages, pronominal systems, or forms of address — code-switching may be accomplished by straightening up one's clothes at the opening of a solemn meeting, or by that brief vocalic exhalation that signals the end of a good laugh.

Underlying language alternation, register shifting, certain types of speech variation, and other communicative behavior is the same process which, I suggest,

ought to be considered code-switching: a switch of frames for interpretation of communicative conduct. Perhaps the subtle mechanisms of most of our daily, monolingual and monolectal code-switching escape us, as speakers and as analysts. We certainly need to refine our tools to examine the emergence, constitution and alternation of communicative codes in context. It is in this direction that I suggest the study of code-switching could proceed.

Notes

[1] I want to thank the participants in John Gumperz's seminar in Conversation Analysis and in Penelope Eckert's course in Language as Community Practice at UC Berkeley in Spring 1990 for discussing this work with me. Special thanks go to Penny Eckert, Susanne Günthner, Marco Jacquemet and Pedro Lewin-Fischer for their insights, and to Al Muth for his careful reading of the paper.

[2] Whenever possible, I will refer only to the original sources, not to subsequent versions of a given work which do not contribute anything substantially different. For example, Heller's 1983 work was published in her edited volume (1988); Calsamiglia and Tusón (1980) was later published in the IJSL in 1984; Woolard (1983a) also appears in Heller (1988); and Gal's article in Heller's volume is basically Gal (1987).

[3] For instance, the role of prosody vis-à-vis language in the signalling of social identity has not been systematically examined. It may well be the case that native intonational patterns are at times more significant than language for invoking group membership.

[4] For Gumperz (1989), contextualization is 'the speakers' and listeners' use of verbal and nonverbal signs to relate what is said at any one time and in any one place to knowledge acquired through past experience, in order to retrieve the presuppositions they must rely on to maintain conversational involvement and assess what is intended' (:1).

[5] Thus, part of the oddity of the comment comes from the journalist's characterization of a sports event in political terms. In the context, humor also stems from the identity of the person being quoted. The ambiguous, love-hate historical relationship between the Galician and Portuguese peoples often shows in mutual negative stereotyping. B's partially unintelligible turn about the 'Portuguese (people)' in line 24 probably reflects one of such stereotypes.

[6] In this work, fonts are used as follows: *Galician*; *'gheada' Galician*; Spanish; Syncretic Galician-Spanish variety. Other transcription conventions include: Pitch and tempo: { } stretch of discourse where phenomenon applies; [ac] accelerated, relatively {[ac] faster } tempo; [hi] higher pitch; [lo] lower pitch. Final tones: ↑ high rising final tone of intonation group; ? high sustained tone; higher pitch over entire group;, mid-sustained tone; / falling tone; // extra-falling tone; - truncated

intonation group. <u>Pausing</u>: .. short pause (less than 0.5 sec. approximately); ... long pause (more than 0.5 sec.). <u>Voice overlapping</u>:

┌beginning and end┐
└of simultaneous t ┘alk. <u>Sound prominence</u>: ' a 'rhythmic or 'emphatic accent; * prominent phrase *accent; **CAPS LOUD**er sounds; [ə] mid centr[ə]l vowel or schwa; : sound l::engthening; - sound interrup- or truncat-; **(word)** reconstructed (s)ound or **(word)**; (=) **(=full form)**; xx unintelligi(xx) syllable; **hh** outbreath**hh,** exhalation; **.h.h** inbreath, **.h.h.h** inhalation; **# # #personal name#** withheld. In the English versions, punctuation marks are used conventionally (e.g. '?' indicates sentence modality, not final tone). Conventions were developed mainly after Gumperz (1982) and Atkinson and Heritage (1984).

[7] In common speech 'to speak with *gheada*' is *ghear*. It can be heard, for example, *Em Cangas todo mundo ghea* 'In Cangas everybody pronounces *gh*'.

[8] Apparent grammatical anomalies in the English version correspond to phenomena found in the original exchange.

[9] The diacritic indicates lexical stress. [espeθjáles] may be either Spanish or dialectal Galician (standard Gal. *especiais*); the final velar nasal in [taŋ] and [eŋ] is characteristic of Galician, and it is transferred to Galiza Spanish; and [pos] may be the monophthongization of either Sp. pues /pwés/ or Gal. *pois* /pójs/.

References

Albert, Ethel M. 1972. Culture patterning of speech behavior in Burundi. Directions in Sociolinguistics, ed. by John J. Gumperz and Dell Hymes, 72-105. New York: Holt, Rinehart and Winston.

Alvarez Cáccamo, Celso. 1989a. Quoting the language(s) of power: voices and code manipulation in reported speech. Paper presented at the 88th Meeting of the American Anthropological Association, Washington, D.C., 1989.
——. 1989b. Variaçom lingüística e o factor social na Galiza. Hispanic Linguistics 2-2.253-98.
——. to appear. Language revival, code manipulation and social power in Galiza. Sociolinguistics of the Spanish-Speaking World, ed. by Carol A. Klee. New York: Bilingual Press / Editorial Bilingüe.

Atkinson, J. Maxwell, and John Heritage (eds.) 1984. Structures of Social Action. Cambridge: Cambridge University Press.

Auer, J. C. P. 1984. Bilingual Conversation. Amsterdam-Philadelphia: John Benjamin Publishing Company.

Blom, Jan-Petter, and John J. Gumperz. 1972. Social meaning in linguistic structure: code-switching in Norway. Directions in Sociolinguistics, ed. by John J. Gumperz and Dell Hymes, 407-434. New York: Holt, Rinehart and Winston.

Bolinger, Dwight. 1985. The inherent iconism of intonation. Iconicity in Syntax, ed. by John Haiman, 97-108. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Brown, Penelope, and Stephen C. Levinson. 1979. Social structure, groups and interaction. Social Markers in Speech, ed. by Klaus R. Scherer and Howard G. Giles, 291-341. Cambridge: Cambridge University Press.

Calsamiglia, Helena, and Empar Tusón. 1980. Us i alternança de llengües en grups de joves d'un barri de Barcelona: Sant Andreu de Palomar. Treballs de Sociolingüística Catalana 3.11-82.

Duranti, Alessandro. to appear. Code-switching and conflict management in a Samoan multiparty interaction. Pacific Studies.

Ervin-Tripp, Susan. 1973. An analysis of the interaction of language, topic and listener. Language Acquisition and Communicative Choice, 239-261. Stanford: Stanford University Press.

Gal, Susan. 1978. Peasant men can't get wives: Language change and sex roles in a bilingual community. LiS 7.1-16.

——. 1979. Language Shift. Social Determinants of Linguistic Change in Bilingual Austria. New York: Academic Press.

——. 1987. Codeswitching and consciousness in the European periphery. American Ethnologist 14-4.637-53.

Gumperz, John J. 1964. Hindi-Punjabi code-switching in Delhi. Proceedings of the Ninth International Congress of Linguists, ed. by H. Lunt, 1115-24. The Hague: Mouton.

——. 1970. Verbal strategies in multilingual communication. Bilingualism and Language Contact, GURT 1970, ed. by James E. Alatis, 129-147. Washington, D.C.: Georgetown University Press.

——. 1982. Discourse Strategies. Cambridge: Cambridge University Press.

——. 1989. Contextualization and understanding. Berkeley Cognitive Science Report no. 59. Institute of Cognitive Science, University of California, Berkeley. To appear in Alessandro Duranti and Charles Goodwin (eds.) Rethinking Context. Cambridge: Cambridge University Press.

Gumperz, John J., and Eleanor Herasimchuk. 1972. The conversational analysis of social meaning: A study of classroom interaction. Sociolinguistics: Current Trends and Prospects, GURT 1972, ed. by Roger W. Shuy, 99-134. Washington, D.C.: Georgetown University Press.

Gumperz, John J., and Eduardo Hernandez-Chavez. 1971. Bilingualism, bidialectalism and classroom interaction. Language in Social Groups, 311-339. Stanford: Stanford University Press.

Heller, Monica S. 1982. Negotiations of language choice in Montreal. Language and Social Identity, ed. by John J. Gumperz, 108-118. Cambridge: Cambridge University Press.

——. 1983. Strategic ambiguity: code-switching in the management of conflict. Paper read at the 12th NWAVE Meetings, Montreal.

—— (ed.) 1988. Codeswitching: Anthropological and Sociolinguistic Perspectives. Berlin: Mouton de Gruyter.

Hill, Jane H. 1985. The grammar of consciousness and the consciousness of grammar. American Ethnologist 12-4.725-37.

Irvine, Judith T. 1979. Formality and informality in communicative events. American Anthropologist 81.773-90.

——. 1982. Language and affect: Some cross-cultural issues. Contemporary Perceptions of Language: Interdisciplinary Dimensions, GURT 1982, ed. by Heidi Byrnes, 31-47. Washington, D.C.: Georgetown University Press.

McClure, Erica, and Malcolm McClure. 1988. Macro- and micro-sociolinguistic dimensions of code-switching in Vingard. Codeswitching: Anthropological and Sociolinguistic Perspectives, ed. by Monica S. Heller, 25-51. Berlin: Mouton de Gruyter.

Mitchell-Kernan, Claudia. 1972. Signifying and marking: two Afro-American speech acts. Directions in Sociolinguistics, ed. by John J. Gumperz and Dell Hymes, 161-79. New York: Holt, Rinehart and Winston.

Ochs, Elinor. to appear. Indexicality and socialization. Cultural Psychology: The Chicago Symposia, ed. by J. Stigler, G. Herd and R. Shweder. Cambridge: Cambridge University Press.

Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. Linguistics 18.581-618.

Scotton, Carol M. 1976. Strategies of neutrality: language choice in uncertain situations. Language 52-4.919-41.
——. 1983. The negotiation of identities in conversation. A theory of markedness and code choice. International Journal of the Sociology of Language 14.115-36.
——. 1988. Codeswitching as indexical of social negotiations. Codeswitching: Anthropological and Sociolinguistic Perspectives, ed. by Monica S. Heller, 151-186. Berlin: Mouton de Gruyter.

Valdés, Guadalupe. 1981. Code-switching as a deliberate verbal strategy: a micro-analysis of direct and indirect requests. Latino Language and Communicative Behavior, ed. by Richard P. Duran, 95-107. Norwood, N.J.: Ablex.

Voloshinov, Valentin N. 1973 [1929]. Marxism and the Philosophy of Language. Cambridge: Harvard University Press.

Weinreich, Uriel. 1953. Languages in Contact. The Hague: Mouton.

Woolard, Kathryn A. 1983a. 'El saben aquel': an example of code manipulation in the course of the telling of a joke. Manuscript. Philadelphia: Graduate School of Education, University of Pennsylvania.
——. 1983b. The Politics of Language and Ethnicity in Barcelona, Spain. Doctoral Dissertation. Berkeley: Department of Anthropology, University of California.

Rules and Representations in Morphology

Stephen R. Anderson*
The Johns Hopkins University

The present paper has an essentially programmatic goal: to lay out some reasons to believe that words ought not to be assigned, as commonly assumed, an internal structure motivated by the morphological relations they bear to other words. That is, I intend to question the idea that words have an 'internal syntax' that consists in a structured arrangement of 'morphemes', and that morphology itself (or at least the part of it that corresponds to 'morphotactics' in structuralist accounts) is essentially the study of this kind of structure.[1] I will suggest that if we construe morphological relations as described by Word Formation rules, the structure of the derivation of a word represents the only sort of information about its relation to other words that can be relevant to any aspect of grammar. Eliminating the assignment of internal morphological structure to words would constitute an advance, because it would render unnecessary a number of otherwise *ad hoc* moves and stipulations at diverse points within current theory whose goal is precisely to ensure that such structure is not referred to. Naturally, it is impossible within the compass of a short paper even to survey all of the ways in which word internal morphological structure has been invoked, let alone to show that all such invocations are dispensable; but it may be possible at least to convince the skeptical reader that the present research program is not a completely implausible one.

We start by asking what information must inescapably be treated as part of the structure of individual words. There seem to be three aspects of a word's representation that have this character: indeed, the essence of being part of a language's lexical stock would seem to be precisely the association of these three sorts of property with one another, as in (1).

(1)   a. Sound: [kʰæt] (with appropriate autosegmental and metrical structure)

      b. Meaning: "A carnivorous mammal (*Felis catus*) long domesticated and kept by man as a pet or for catching rats and mice"

      c. Syntax: [+Noun] (or something)

A word is an association between sound and meaning that fills some syntactic function, so this much is irreducible. The question to be considered below is whether words should in general be assigned any other sorts of internal structure. It is of course quite clear and uncontroversial that we know many things about words beyond the basic information in (1), but this fact alone does not necessarily justify the attribution of further structure to words if there are other ways of accounting for the kinds of 'knowledge' involved.

Since structuralist times, it has been assumed that (independent of their phonological organization) words are divisible into **morphemes**. A theory gradually grew up according to which the phonological forms of these elements are separated from one another by quasi-phonological 'boundaries', and organized into a hierarchical structure like that of the Phrase-Marker in (2).

(2)

'discontentedness'
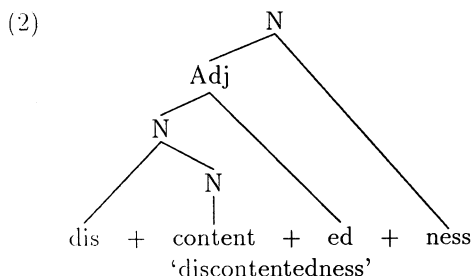
Variations on this scenario are represented, for instance, by the views of Selkirk 1982, Di Sciullo and Williams 1987, and Halle 1990, all of whom treat morphology as in one way or another the 'internal syntax' of words.

We can contrast such views, based on morphemes, with a picture of morphology as based on a system of rules that map words (or stems) onto other words. There are in fact a number of morphological theories presently on the market based centrally on Word Formation Rules rather than on morphemes of the more traditional sort, but to be more concrete we adopt the specific proposals of Anderson forthcoming. Let us call this the 'A-Morphous' view of morphology. On this picture, the structure of *discontentedness* is given by a derivation like that in (3).

(3) $[_N$ content$] \xrightarrow{\mathcal{R}_{dis}} [_N$ discontent$] \xrightarrow{\mathcal{R}_{-ed}} [_{Adj}$ discontented$] \xrightarrow{\mathcal{R}_{-ness}} [_N$ discontentedness$]$

The individual rules of a language's morphology describe the (more or less) systematic relations that obtain among classes of words. Each step of a derivation like (3) invokes such a rule, which is a mapping between the phonology, the semantics, and the syntax of one such class (its 'inputs') onto the properties of another class (its 'outputs'). The derivation thus represents the place of an individual word in a network of such relations that constitutes the lexicon (in at least one reasonable, pretheoretic sense of the term) of a language. It expresses the same facts about subparts of words that are themselves words, relative scope of morphological operations, etc. as the Phrase-Marker, but without imposing a distinct structure on derived words themselves.

If a structure such as that in (2) has to be imposed on words, that would presumably constitute an argument for the reality of the constituents of such

structures: morphemes. On the other hand, a theory (such as that of Anderson 1988, forthcoming) which dispenses with morphemes as units has the consequence that such structure will only arise when explicitly stipulated in the structural change of a morphological operation. In the usual case, 'complex' words will have only phonological and semantic structure, together with syntactic properties.

Now while no one would contest the suggestion that information such as that presented in (2) is linguistically 'real', the observation that much the same information is presented in (3) should make it clear that there are real issues concerning the way this information should be represented. Do words have a morphological **content**, as assumed in (2), in addition to their phonological, semantic and syntactic content? If so, is the attribution of such internal structure the best way to represent that content? In that case, we ought to find that the structure attributed to words is actually referenced by rules of the grammar that operate on words. On the other hand, if the relation of words to other words is represented by their derivation as in (3), not by shared internal structure as implied by morpheme based views, we ought to find that morphological structure has a much narrower relevance. The choice between (2) and (3) is thus an empirical one, at least in principle.

So what sorts of motivation might we find for a structure like that in (2)? In theory at least, such motivation could come from any of a number of places. We might, for example, have *a priori* reasons to believe in it: it might be that the independent principles of $\overline{X}$-structure extend directly to describe the internal nature of words. This, however, seems quite unlikely. Zwicky 1990 has recently summarized a dozen or so ways in which word-internal structure differs significantly from syntactic structure. In addition to Zwicky's points, Williams 1989 observes a number of major differences that follow from the absence of a word-internal analog of the syntactic notion of 'Maximal projection'. Although Williams' goal is ultimately to justify a relation between morphological and syntactic structure, the effect of his argument is virtually to eliminate any significant resemblance between the two. The similarities that remain between the two sorts of structure are extremely limited. Even the claim that "both use concatenation as the basic operation" (hardly a profound or substantive theoretical resemblance between two systems that have been formulated as concatenative algebras!) may well not be true in all cases, once one takes seriously the range of morphological processes in various languages that cannot in fact be described adequately as simple affixation. The end result of the observations of Zwicky, Williams and others is to eliminate any claim that syntactic $\overline{X}$-structure itself motivates word-internal structure like that in (2).

Absent any such *a priori* justification, evidence for the structure in (2) would be provided by rules of any component of grammar that necessarily

referred to such structure. For instance, semantics could provide the motivation we seek if we believed that it was necessary to invoke morphological structure *per se* in the computation of the meaning of a word. Among other things, (2) represents relations of relative semantic scope; but the same scope relations that (2) indicates are also indicated in a derivation like (3). There seems no reason to doubt that the relevant scope relations can be quite adequately represented by treating the semantic aspect of Word Formation Rules as operating on the basis of the semantic content of the word that constitutes the word's input, and thus the position of a particular rule in the derivation reflects the semantic material that falls within its scope without requiring us to assign further structure for this purpose.

If reorganization of a word's internal structure were a crucial part of the mapping from sentences onto their meanings (as proposed, for example, by Pesetsky 1985), this would of course constitute an argument. Pesetsky's account of 'bracketing paradoxes,' however, has spawned a virtually limitless range of alternative proposals, not in general requiring readjustment of a structure such as that in (2) and so not providing motivation for it. We do not summarize these alternatives here, but simply note that others have (for completely unrelated reasons) proposed ways of describing mismatches between apparent phonological and semantic structure of the sort described by Pesetsky without relying on the sort of structure we are interested in here.

The syntax could provide the required motivation if it were allowed to examine or manipulate the internal forms of words. This is directly contrary to the most natural construal of the Lexicalist Hypothesis, however. In fact, it can be shown that representations such as (2) are both too weak (i.e., they provide too little information) and too strong (they provide too much information of an inappropriate sort) to support the informational interchange that exists between words and syntax. An alternative view is presented in Anderson (forthcoming) which eliminates this sort of motivation for assigning internal morphological structure to words. On that basis, the denial that there is any structure like that in (2) provides an excellent way of eliminating the stipulative character of the claim that the syntax neither sees nor manipulates it (the essential content of the Lexicalist Hypothesis).

The phonology could provide such motivation if it were allowed to refer to the presence of internal structure. Of course, the 'Standard Theory' of Chomsky and Halle 1968 assumed that all sorts of structure was indeed available, but the past twenty years or so have been marked by a continuing retreat from that position. Conditions such as 'Bracket Erasure,' 'Tier Conflation' and the like are intended to restrict the phonology to as little structural information as possible. In fact, having Bracket Erasure (or its analog) apply at the end of each cycle, as is widely assumed, limits the information available to the phonology to precisely the content of the current morphological operation defining

that cycle—and that same information can be accessed via an (independently necessary) condition on rules that only apply in derived environments without requiring that the structure itself be accessible to phonological rules.

It looks as if evidence for a structure like that in (2), if it exists, would have to come from the morphology itself. We might find evidence for the presence of structure if we had a rule that, say, inserted affixal material precisely between two morphological units. A rule might infix a particular marker precisely before the final suffix (or after the initial prefix) present in its input form. Such rules do not appear to exist, however: rules of infixation seem always to place their affix with regard to material that can be characterized in phonological terms. Infixes thus appear after (or before) a single consonant, a syllable or syllable nucleus, a prosodic foot, etc., but never after (or before) a single affix regardless of its phonological shape. This fact suggests that information about the morphological (as opposed to phonological) constituency of words is not available to morphological rules.

As in the syntactic domain, the phonological case provides a genuine theoretical advantage that accrues to non structure-building views of morphology. Theories that posit the erasure of brackets, or the conflation of tiers, etc., can avoid stipulating any such operations (and *a fortiori* avoid ordering them at some particular point) if in fact no brackets or tiers (both ways of indicating morphological structure of the sort that appears in (2)) were introduced in the first place. In the one place where reference to structure appears motivated, it can be replaced with a condition on the way in which (at least some) rules operate—the 'derived environment' condition.

In addition to the general points made above, however, there are several other proposals that have been made in the literature that would indeed require morphological operations to have access to information like that in (2). For example, Aronoff 1976 proposes that rules of **truncation** operate precisely on specified affixes, whose identification within a complex word would thus be necessary. According to Aronoff, the affix -*ate* is deleted before some other affixes, such as -*able*, in words like *demonstrable* (from *demonstrate*). Since instances of the sequence -*ate* which are not affixes are not truncated (cf. *debatable*, not *debable, from *debate*), he argues that it is impossible to replace reference to the affix by a purely phonological description. Truncation rules would thus require information about the morphological structure of words which is not purely phonological.

The argument that truncation here is based on morphological and not phonological structure is not particularly secure, however. The absence of forms like *debable, *inflable, etc. would also follow from from a requirement that the truncation in question applies only to unstressed instances of -*ate*. Furthermore, this affix appears in two classes of words, associated with either 'level 1' or 'level 2' phonology.[2] The 'level 1' forms undergo truncation of -*ate*,

but the 'level 2' forms do not: see pairs such as *demónstrable* (level 1) vs. *demonstrátable* (level 2). While the (non-truncating) level 2 formation seems quite productive, the (truncating) level 1 formation is, as we would expect, somewhat less so. As a result, even in cases where we have no particular reason to doubt the affixal status of *-ate*, the only available form in *-able* may not show truncation: consider the word *truncate* itself, from which we find *truncatable* but not **truncable*. The absence of e.g. **translable* might thus just be a lexical gap.

Furthermore, other instances of truncation seem clearly to involve morphological non-constituents. A considerable range of such forms in French is provided by Corbin 1987, p. 345, including the following:

(4)   a. virus – viral; cactus – cactée; rectum – rectal; tétanos – tétanique

     b. liquide – liquéfier – liqueur; stupide – stupéfier – stupeur *vs.*

     c. rigide – rigidifier – rigueur; humide – humidifier – humeur

     d. certain – certitude; caillou – caillasse

     e. charité – charitable; hérédité – héréditaire; vanité – vaniteux

     f. adroit – adresse, maladroit – maladresse

In none of these cases have we any reason (apart from truncation) to believe that the truncated material is a morphological unit in French. We conclude that rules of truncation do not provide clear evidence for imposing morphological structure on words.

When we ask why languages should contain rules of truncation at all (assuming they do: see Kiparsky 1982 for a contrary view), it seems rather unlikely that they should have the morphologically sensitive character originally attributed to them by Aronoff. The typical case in which one is tempted to posit truncation, as in the cases cited above in (4), is when a language (e.g. French) has borrowed (or even inherited) a large number of formations from another (e.g. Greek or Latin) that are built on the same stem, but without borrowing (or inheriting) the entire morphology of the source language—particularly its inflectional pattern. If the source language is one in which all (or at least most) surface forms of words contain an inflectional suffix, then the relation between a basic word and a derivational formation from the same stem will appear to consist in the replacement of one suffix (an inflectional one, or perhaps a derivational one corresponding to a non-borrowed formation) with another (the derivational suffix). Insofar as some of the morphological material in the original language is not part of the system of the borrowing language, an apparent truncation applies. Truncation is thus an attempt (at least in some cases) to cope precisely with the replacement of material that is *not* a morphological unit in the 'truncating' language. It would be remarkable, in

that event, if it were constrained to apply only to material that *did* constitute a morphological unit.

Another morphologically based argument is offered by Fabb 1988, who proposes explicitly that "all internal brackets are visible to all suffixes" (p. 533). His justification for this is the claim that English exhibits a number of suffixes that are constrained not to be added to already-suffixed words—a requirement that obviously could not be met unless internally complex words could be distinguished from others. This argument suffers from serious empirical problems, however. In fact, a quick check of such reference works as Marchand 1969 and Jespersen 1942 turns up examples of already suffixed words that can be further suffixed in virtually all of Fabb's affix classes. Indeed, he himself gives as an example at least one form of this sort: *robbery*, presented as an instance of the formation of Nouns from Noun stems by the suffixation of −*y* is surely from *robber*, itself a suffixed form derived from *rob*. Some other examples of already suffixed bases in the derivational categories discussed by Fabb include:

(5)   a. N+age]$_N$: brokerage, porterage, readerage, percentage,...

      b. V+ant]$_{Adj}$: significant; also radiant, stimulant, etc. with truncation of -ate;

      c. N+ful]$_{Adj}$: healthful, truthful, meaningful

      d. N+ism]$_N$: gangsterism, behaviorism, favoritism, aristotelianism,...

      e. N+ize]$_V$: computerize, puritanize, italianize, signalize,...

      f. N,Adj+ly]$_{Adj}$: maidenly, scholarly, loverly, northerly, westerly

      g. V+ment]$_N$: chastisement, aggrandizement, betterment, enlightenment,...

      h. N+ous]$_{Adj}$: felonious, erroneous, disputatious, flirtatious

All of Fabb's affixes also appear on the second elements of compounds (with scope over the entire compound) and on already prefixed words, even those that are otherwise not widely attested with other affixes.

Fabb's methodology seems to have involved extensive reliance on listing in Walker 1924 as a criterion for the possibility of a formation. As argued extensively by Corbin (*op. cit.*), dictionary listing is far from being an appropriate measure of the potential domain and range of word formational processes, and I can only conclude that no argument of the sort Fabb proposes has actually been made. Indeed, the kind of selection he proposes would be quite unusual: normally subcategorization is for some positive property of the subcategorized element rather than for a purely negative aspect of its structure.

A final argument known to me from this general class is made by Halle in various recent works dealing with morphological theory. Since much of this

discussion remains unpublished, I must ask the reader to accept my presentation of Halle's position. The issue concerns the prefixes[3] that mark the person of arguments of the Verb in Georgian, about which the basic observation is that when two such prefixes are called for, only one appears:

(6) a. mo-v-ḳlav 'I will kill him'

b. mo-g-ḳlav-s 'he will kill you'

c. mo-ḳlav-s 'he will kill him'

d. mo-g-ḳlav 'I will kill you' (*mo-v-g-ḳlav, *mo-g-v-ḳlav)

Halle has proposed to account for this by having first a set of rules to develop Object markers (such as /g–/ '2nd person object') and then a set of rules to develop Subject markers (such as /v–/ '1st person Subject'). The complementarity observed above is then claimed to follow from a restriction that the subject marking rules apply only to un-prefixed forms. It appears that if this is to work, enough structure must be visible to distinguish prefixed from unprefixed forms.

Halle appears to support the plausibility of this analysis by an appeal to the fact that in German, the prefix /ge–/ 'past participle' is not added to verbs that already bear a prefix. As Kiparsky 1966 showed long ago, not only verbs with prefixes (e.g. *besprechen* 'discuss') but also those with exceptional non-initial stress (e.g., *riskieren* 'risk') have participles without /ge–/ (*besprochen* and *riskiert* respectively).

If one wished to maintain the claim that it was the prefixed (as opposed to unprefixed) status of forms like *besprechen* that blocks the addition of /ge–/, it might be suggested that prefixes like /ge–/ can only be added to forms that are not only unprefixed, but also bear some feature such as [+Native]. Since *riskieren* is undoubtedly [−Native] (as shown by the suffix *–ieren*), it would fail to take /ge–/ as well. For this analysis to be taken seriously, however, we must first ask what 'non-native' and 'prefixed' stems have in common, and at least at first glance, it appears that the answer is that these are the cases in which main stress does not fall on the initial syllable. The disjunctive statement seems to miss the underlying regularity. More importantly, though, we should ask what the content is of the proposed feature [±Native]. Does it, in particular, predict any aspect of the behavior of [−Native] words other than their non-initial stress? If so, it is not evident what that aspect would be. But if the content of '[−Native]' is simply 'non-initial stress' this analysis differs from the one based on the location of stress (which makes no reference to 'prefixed' vs. 'unprefixed' status) only in saying twice what need only be said once, since it is also a property of the inseparable prefixes that they cannot bear stress.

If the class of [−Native] words is defined in some way other than as 'non-initially stressed', it ought to be possible to find cases of non-initially stressed [+Native] words, and of initially stressed [−Native] ones. We ought to expect the former to take prefixes like /ge−/ and the latter to reject them. The difficulty of confirming independently the status of proposed examples of the first type as properly [+Native] makes it difficult to propose candidates which might either confirm or to refute the suggestion. But of the second sort, there are examples. The verb *boxen*, 'to box' for instance, is apparently [−Native]. Similarly, *toasten* 'to propose or drink a toast to' is apparently felt as non-native, as confirmed partly by its spelling and partly by the Viennese dialect form *toastieren*. It seems likely that *leasen* is also felt synchronically as non-native, though apart from the spelling this is again difficult to confirm. All of these, being monosyllabic stems, bear initial stress despite their apparent status as [−Native]—and all of them accept /ge−/ happily in their past participles (*geboxt, getoastet, geleast*).[4]

It looks, therefore, as if in order to maintain the suggestion that prefixes like /ge−/ are added exactly to unprefixed native stems, it will be necessary to circumscribe the class of non-native stems in a less than intuitive fashion—essentially reducing [−Native] to a diacritic for 'non-initially stressed'. But in that case, surely, the analysis based on stress is preferable, and so German does not serve as an example of a language with a restriction of certain processes to 'unprefixed' stems. The correct generalization makes no reference to internal morphological structure: it is simply that /ge−/ is only attached to stems that bear initial stress.

Regardless of what is going on in German, though, Halle's account would be quite interesting if it went through for Georgian. The proposed condition, however, is simply incorrect. We can ignore the fact that Georgian verbs can also be prefixed with an aspectual preverb (e.g., the /mo−/ in the forms above, rather parallel to the corresponding Russian preverbs), where the presence of such a preverb has no effect on whether or not /v−/ can be prefixed. Besides this, Georgian Verbs can contain other prefixes: there is an important class, called the 'pre-radical vowels', which occur for a variety of reasons, derivational and inflectional. These are surely prefixes, but do not block the appearance of /v−/.

Furthermore, there are other prefixes that mark indirect objects. For the first and second persons, these are the same as those marking direct objects, but for third person indirect objects the situation is more complicated. When the pre-radical vowel appearing before a given verb stem is /a−/ or /e−/, the third person indirect object is unmarked. When it is /i−/, third person indirect objects are marked by changing this /i−/ to /u−/. But when there is no preradical vowel, a consonantal prefix marks third person indirect objects, at least in the 'standard' dialect. This is the reflex of Old Georgian /x−/, and

appears in the modern language as /s–/ before coronal non-continuants, /h–/ before labial, velar and uvular stops, and as ∅ elsewhere. Now this prefix has a rather peculiar status: generally its appearance is blocked by the appearance of another prefix (thus, *v-txov* 'I ask him it', where the only prefix is that marking first person subject). There is some fluctuation, however, and forms such as *v-s-txov* 'I ask him it' (where the indirect object prefix appears as well) also occur.[5] This state of affairs can be analyzed as a fluctuation in whether the s/h/∅ prefix is associated with the other verb person prefixes (and thus mutually exclusive with the other person prefixes, giving forms like *v-txov*), or with the pre-radical vowels (in which case its appearance is compatible with other prefixes, giving forms like *v-s-txov*). On Halle's view, it is not clear how the form *v-s-txov* could be derived at all, since the /v–/ here appears in an evidently 'prefixed' form. Notice, by the way, that there is absolutely no variation in the case of the other prefixes: *g-ḳlav* is the only possible form for 'I am killing you', and **v-g-ḳlav* is completely impossible.

The generalization, then, is not that /v–/ cannot be added to prefixed forms. In fact, aspectual prefixes, preradical vowels, and the person marker s/h/∅ (for some speakers) can all be prefixed to verbs without blocking the introduction of /v–/. Rather, it is the *specific* prefix /g–/ (and by implication, perhaps, also /m–/ '1st person singular object' and /gv–/ '1st person plural object', though these could never cooccur with first person subjects for syntactic reasons) whose incompatibility with /v–/ must be described. Apparently, then, the condition on the /v–/ rule must be "does not apply if the /g–/ rule has already applied", just as the condition on the s/h/∅ rule, for those speakers who assign it to the verb person prefix class, is "does not apply if the /v–/ rule[6] has already applied."

In these instances, the condition is thus a matter of the mutually exclusive character of certain specific elements, not a limitation (such as "does not contain a prefix") on the formal makeup of the stem to which a morphological element is to be added. But wait: isn't that exactly the sort of thing I am supposed to be arguing does not occur in grammatical systems? Surely if the account offered above is correct, the grammar of Georgian must be able to identify specific prefixes, a task for which the kind of structure that appears in (2) would appear to be well suited.

The solution to this puzzle is to treat the relation of exclusiveness between the two prefixes as a matter of disjunction between two rules, rather than as a subcategorization restriction on one of the affixes. A stipulation that two (adjacent) rules are disjunctive in this fashion is quite comparable (though not equivalent) to the more traditional claim that the rules in question belong to the same position class. Such an account was proposed in Anderson 1986, and Halle's argument was intended precisely to show that the kind of disjunction claimed there to exist can be replaced with positive conditions on individual

rules (much like those offered by Fabb). But if we recognize that the complementarity observed in the Georgian prefixes is a relation between Word Formation Rules, we do not need to assign internal structure to the forms themselves in order to account for the facts. Such a condition on representations can (and apparently must) be replaced by a condition on the interaction of rules within a derivation.

The point of this discussion is that much of what might be done in the structure-based account by various sorts of subcategorization restriction (assuming that conditions such as "only attaches to morphologically simplex forms" could be stated at all) can be formulated in the structureless theory as matters of the internal organization of a system of rules. Now we clearly need some such organization in any event: relative order of affixes, for example, is reconstructed by the relative order of rules in the sort of system I am proposing. But if we assumed that morphological structure is recorded whenever we perform morphological operations, and (what is more important) that that structure can be referred to at various points later in the derivation, we would be able to formulate essentially global conditions. Such conditions, making essential reference to arbitrary earlier stages of a derivation, cannot be expressed as conditions on adjacent steps (the conceptual content of ordering restrictions on pairs of rules). There is thus a substantial theoretical advantage to be achieved by pursuing a program on which morphological form is represented as the application of rules, and abandoning the morphological analogs of the Phrase Markers that are so important to syntactic form. We also, of course, avoid confusing claims about the interrelations of words within the lexicon of a language with claims about their internal structure and the information they bear. And we see that one more potential source of support for the notion that words are built up as structured concatenations of minimal signs ("morphemes") falls by the wayside.

Footnotes

[1]It should be made clear at the outset that the goal of eliminating word internal morphological structure which I pursue here is not intended to be applicable to the analysis of compounds, but only to the traditional categories of 'derivation' and 'inflection'. Even within these latter domains, examples exist in which internal structure must be presumed to exist: for an instance, see the category of 'middle' verbs with the ending –*st* in Icelandic (Anderson 1990).

I do not mean to suggest that Word Formation Rules *cannot* build internal structure, but only that when they do so, this must be treated as a specifically stipulated aspect of their structural change, and not as the default consequence of any morphological operation.

[2]This distinction is couched here in the terms of Lexical Phonology, but is also familiar as the difference between the phonology of '+ boundary' and '# boundary' affixes, etc. Nothing hinges on the choice of terminology here.

[3]A corresponding argument can be made in connection with Verbal suffixes for person and number, but is not given here for reasons of time.

[4]I am indebted to Gert Webelhuth for helpful discussion of these forms.

[5]See Vamling 1989 for some discussion.

[6]Again for syntactic reasons, this is the only other verbal agreement prefix that could appear in a form with a third person IO that must be marked by s/h/∅.

References

Anderson, Stephen R. 1986. Disjunctive ordering in inflectional morphology. *Natural Language and Linguistic Theory* 4:1–32.

Anderson, Stephen R. 1988. Morphological theory. In Frederick J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey*, Vol. I, 146–191. Cambridge: Cambridge University Press.

Anderson, Stephen R. 1989. On looking into words. Paper read at the 6th meeting of the Eastern States Conference on Linguistics, University of Delaware, 7 October. To appear in the proceedings of the conference.

Anderson, Stephen R. 1990. The grammar of Icelandic verbs in *−st*. In Joan Maling and Annie Zaenen (Eds.), *Icelandic Syntax*, Vol. 23 of *Syntax & Semantics*. New York: Academic Press.

Anderson, Stephen R. forthcoming. *A–Morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Corbin, Danielle. 1987. *Morphologie dérivationelle et structuration du lexique*. Vol. 193, 194 of *Linguistische Arbeiten*. Tübingen: Max Niemeyer Verlag. [two volumes].

Di Sciullo, Anna-Maria, and Edwin Williams. 1987. *On the Definition of Word.* Cambridge MA: M. I. T. Press.

Fabb, Nigel. 1988. English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory* 6:527–539.

Halle, Morris. 1990. An approach to morphology. *Proceedings of the North Eastern Linguistic Society* 20.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles.* Vol. VI. Copenhagen: Ejnar Munksgaard.

Kiparsky, Paul. 1966. über den deutschen Akzent. *Studia Grammatica* 7:69–98.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In *Linguistics in the Morning Calm*, 3–91. Seoul: Hanshin Publishing Company.

Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word Formation.* München: Beck. 2nd edition.

Pesetsky, David. 1985. Morphology and logical form. *Linguistic Inquiry* 16:193–246.

Selkirk, Elizabeth. 1982. *The Syntax of Words.* Cambridge MA: M.I.T. Press.

Vamling, Karina. 1989. *Complementation in Georgian.* Lund: Lund University Press.

Walker, John. 1924. *The Rhyming Dictionary.* London: Routledge.

Williams, Edwin S. 1989. Maximal projections in words and phrases. In Mark R. Baltin and Anthony S. Kroch (Eds.), *Alternative Conceptions of Phrase Structure*, 280–291. Chicago: University of Chicago Press.

Zwicky, Arnold M. 1990. Some choices in the theory of morphology. In *Formal Grammar: Theory and Implementation*, Vancouver Studies in Cognitive Science. Vancouver: University of British Columbia. [to appear].

DISEMBODIED RULES VERSUS PATTERNS IN THE LEXICON
Testing the Psychological Reality of Spanish Stress Rules[*]

Jon Aske
University of California, Berkeley

"However, we are not interested in just *any* kind of linguistic description that is compatible with data. We are concerned about that specific way (or those ways) of organizing linguistic knowledge that speakers actually use."
–Per Linell (1979:18)

"We believe that the results of these experiments suggest ... that the speakers make reference to the words in the lexicon, not to lexicon-independent rules, when making judgments of the kind we required of them."
–John Ohala & Manjari Ohala (1986:248)

This paper reports on an experiment carried out to test, or at least gain some insight into, the psychological reality of proposed rules for Spanish (non-verb) stress. Although talk of psychological reality seems to have fallen out of fashion in the 1980s, the issue is still relevant and important--not just "self-indulgent methodological agonizing," as one critic has referred to it--and much insight can be gained into the nature of language and its regularities by means of experimentation, an experiment being nothing other than an "observation under carefully controlled conditions" (Ohala 1987:207; cf. Ewen & Anderson 1986, Ohala & Jaeger 1986). The results of this experiment suggest that the stress of Spanish nouns and adjectives is not stored in the form of rules, in spite of being about 95% predictable. That is, speakers do not seem to make abstracted, or 'disembodied' generalizations about the regularities involved. This suggests that, although the linguist's rules are valid generalizations about the surface patterns or output, they aren't necessarily a good model of the internal 'mechanism' producing that output.

1. Rules and psychological reality

Although the use of rules as a device to represent phonological regularities did not start with Generative Phonology,[1] rules did take on a new dimension with the advent of this school of phonological analysis. They had a very central role in generative grammar from the beginning, both in syntax and in phonology,[2] and became the single mechanism to account for every ounce of regularity or redundancy, from low level, phonetic/allophonic variation to morphophonemic alternations. A consequence of this emphasis on rules to capture generalizations and regularities about the 'surface' phonological shape of forms was, among other things, "a change in phonology in the direction of much more abstract representations than those permitted within a theory which concentrated on biunique phonemics" (Anderson 1985:321). And the more abstract the representations became and the larger the number of regularities and subregularities to be explained by rules, the less regular these became, that is, the more they had to be supplemented by additional conditions on the application of the rules, leading to quite abstract representations.

The emphasis on rules in Generative Grammar was accompanied by a cognitive ("mentalist") framework which attributed psychological reality to the linguist's rules and other constructs. Thus, the individual speakers learning their language supposedly extracted the same generalizations as the linguist from the input utterances to which they were exposed, and the regularities obtained were stored in the form of rules into which abstract, fully idiosyncratic representations were 'fed' to produce the surface patterns of the language. Both the centrality of the interest in the psychological reality of linguistic description and the use of systems of rules to formalize these descriptions was quite revolutionary at the time. However, it was never quite clear exactly what the mental equivalents of the rules

and representations were to be. Soon it became clear, however, that the **process view** of rules, though perhaps plausible for low level phonetic alternations,[3] would not work for the increasingly abstract rules devised by generativists, nor for other constructs such as features,[4] rule ordering, and the cycle. The regularities that generative linguists in the 1960s extracted from the lexicon became increasingly less like anything speakers are likely to use or possess and this brought a reaction in the field, a 'psychological reality' backlash (cf. e.g. Kiparsky 1968, Derwing 1973, Hooper 1976).

Many linguists, wanting to 'have their cake and eat it too,' have adopted what Per Linell (1979) calls a "weak representationalist position," according to which not "every aspect or detail of the theory is assumed to be isomorphic to some psychological (or neurological) counterpart" (p. 11), but rather "the relationship between the theoretical grammar model and the speaker's internalized knowledge is more indirect" and "only *some* aspects of [the] theory mirror psychological structures" (p. 11). The weakest claim along these lines is that "only the *output* of the grammar relates to psychological reality, i.e. the strings generated are those which are judged as 'correct' by the speakers" (p. 12). Thus we should continue to do what linguists have always done and are good at, namely, look for and extract patterns and redundancies from the output/corpus. This attitude, according to Linell, is not really mentalistic at all, despite claims to the contrary, but rather merely formalistic.[5]

This is not to say that all modern phonologists would be in agreement with the cognitive framework or, if so, to similar degrees. Linell (1979) has a useful typology of linguists according to where they stand on the psychological reality issue. First, there are the **radical physicalists**, or radical behaviorists, for whom talk about the psychological reality of abstract concepts is nonsense. Next there are the **pessimists**, according to whom "[s]peakers are assumed to possess knowledge of their language, i.e. a grammar with a more or less specific organization which enables them to use their language correctly, but it is considered to be an unattainable goal to find out what properties these psychological structures have. Therefore, linguists should avoid these problems and do 'autonomous linguistics', i.e. establish linguistic generalizations by purely structural methods" (p. 4). As Linell observes, this is a respectable position, though not justified in our view. For another set of linguists, the **naive optimists**, mostly Chomsky and his followers, "speakers are assumed to have highly integrated and interindividually similar 'mental grammars'" (p. 5) organized around symbol manipulation. For them, "since many irrelevant factors intervene in performance the best way to determine mental grammars would then be to apply formal-linguistic methods in trying to investigate general and abstract conditions on linguistic structures" (p. 5).[6] According to Linell, "[i]n many respects, this kind of linguistics is an extreme form of structuralism or of 'autonomous linguistics,' i.e. a linguistics which refuses to utilize external evidence" (p. 5). Finally, the 'good guys' in Linell's typology are what he calls **moderate realists**. For them "speakers are assumed to have organized knowledge of their language in some specific ways," ways which can and must be sought by the investigator. But, he adds, "an investigation of psychological realities cannot be pursued with purely linguistic-structural methods. Instead, many types of 'external' evidence must be exploited... Also, one needs plausible 'metaphysical' assumptions about the nature of language, language acquisition and use, the properties of the mind, etc." (p. 5). It is in keeping with this approach that I will investigate the psychological implementation of Spanish stress patterns in this paper.

2. Some generativist rule accounts of Spanish stress

Before generativist accounts of Spanish stress, it was fairly uncontroversial that this stress was phonemic, and thus listed in the lexicon, since pairs and triplets of words that differ only in their stress pattern are not uncommon, e.g. *intérprete*, '*n.* interpreter,'

interpréte, 'v. I/he/she interpret' (present subjunctive), *interpreté*, 'v. I interpreted' (preterit indicative), or *bostézo*, 'n. yawn', 'v. I yawn,' and *bostezó*, 'v. s/he yawned.' Generative phonologists, on the other hand, emphasized the great regularity and predictability of Spanish stress (as opposed to its partial idiosyncracy and occasional meaningfulness), especially in non-verb forms. Doublets and triplets always involve at most one noun and from one to three verb forms. Thus if verb and non-verb stress are accounted for differently, i.e. with different rules, one can very well say that stress in Spanish, at least non-verb stress, is quite predictable.

This yearning to see regularity everywhere led early generative phonologists working on Spanish to find (phonological) regularity in even the five percent of Spanish words with irregular stress, and, furthermore, to argue that there are "essentially valid arguments for stress assignment by the Latin stress rule," a rule "which, roughly, assigns stress to the penultimate syllable of polysyllabic words if that syllable is 'strong' (contains a tense vowel, or a lax vowel followed by two or more consonants) and to the antepenultimate syllable if the penultimate is 'weak' (contains a lax vowel followed by at most one consonant)" (Harris 1969:118; cf. (1) below). Never mind that Spanish shows no evidence of tense vs. lax vowels, other than perhaps some frozen reflexes, and that many more Spanish words are oxytonic than proparoxytonic; these 'minor' discrepancies could always be accounted for ("explained") by appealing to abstract segments (final vowels for words ending in a consonant, the vast majority of the 'surface' oxytonic words) and abstract features (tense vs. lax vowels). (For verb forms the story repeats itself, with the use of mechanisms that either mirror diachrony or are entirely *ad hoc*.)[7]

(1) $\text{V} \rightarrow \text{[1stress]} / \underline{} \begin{Bmatrix} (\ C_0\ (\ \check{V}\ C_0^1\ (L))\ V\ )\ C_0\ \#]_{N,A} \\ (\ ([\text{-perf}])\ C_0\ V)\ C_0\ \#]_V \end{Bmatrix}$

The 1970s saw some complaints about Harris' fantastic accounts of Spanish stress. Stanley Whitley (1976), for instance, gave an insider's critique in which he faults Harris 1969 with overzealousness and argues that Spanish stress is distinctive, that Harris' account is excessively abstract (though Whitley's account is still rule oriented and not devoid of abstraction), and that, although it "appears to effect extensive savings in underlying representation,... the net effect is actually one of complication and a corresponding loss of adequacy and predictiveness" (p. 301).

Joan Hooper & Tracy Terrell (1976) put forth an account of Spanish stress from a Natural Generative Phonology perspective (cf. Hooper 1976), an attempt to put constraints on abstractness by specifying that (1) "the rules of the grammar are not extrinsically ordered," and that (2) "underlying phonological contrasts are limited to direct manifestations of surface forms." Thus, "the result is a grammar in which all the rules express true generalizations about surface forms" (p. 65). They espouse a separation between phonological rules (P-rules), or those that only have phonetic information in them, and morphophonological rules (MP-rules), which incorporate grammatical as well as phonological information.[8] Since, obviously, the Spanish stress rules do not qualify as P-rules (i.e. they are different for verb forms than for nouns and adjectives), they argue that "Spanish stress [is] morphologically determined in relation to the stem of the word" (p. 64). Concentrating on non-verb stress,[9] they show relative frequencies of patterns from a sampling from a dictionary, which clearly reveal what everybody knew, namely that the true ('surface') generalization about Spanish stress is that words that end in a consonant are oxytonic, whereas those that end in a vowel (*-V*) or plural *-s*[10] are paroxytonic:

(2) a. $[... \text{V}\ C_0\ \text{V}\ (\text{s})\ \#]_{N,A}$ Penultimate stress: ~95%; Final stress: ~4%
   b. $[... \text{V}\ \text{C}\ \#]_{N,A}$ Penultimate stress: ~4%; Final stress: ~95%

Hooper & Terrell's morphological solution for non-verb stress takes the final vowel in (2a)

type forms to be a thematic vowel, i.e. a vowel outside the stem (/a/, /o/, or /e/), e.g. *ventán-a(-s)*, 'window(s)', which is missing from the (2b) type forms, e.g. *jabón*, 'soap'. Thus the MP-rule which accounts for the stress of some 95% of Spanish words is the following:

(3) $V \rightarrow$ [+stress] / ___ C_0]STEM

In other words, the 'inflectional' vowel, if there is one, is ignored. This takes care of the vast majority of Spanish words, and the exceptions, which will be marked in the lexicon, are reduced to the following for Hooper & Terrell:

1. Paroxytonic words ending in a consonant (i.e. without a 'thematic' or 'class' vowel: (a) about 20 words ending in /-r/, e.g. *cadáver*, 'corpse', *ámbar*, 'amber'; (b) about 15 words ending in /-il/, e.g. *fácil*, 'easy'; (c) about 10 words ending in /-en/, e.g. *crímen*, 'crime'; (d) 3 words in {-z} (= /θ/ or /s/, depending on the dialect), e.g. *lápiz*, 'pencil'; and (e) 2 words in /-d/ (pronounced [d], [t], [θ], or zero, depending on the dialect), cf. *césped*, 'lawn', and *huésped*, 'guest'. All these words are 'accounted for' by means of a diacritic mark/feature on the final vowel.
2. Oxytonic words ending in a vowel: *-í* (7 ea.), *-ú* (5 ea.), *-á* (4 ea.), e.g. *mamá*, 'mom', *-ó* (2 ea.), e.g. *dominó*, 'domino', and *-é* (10 ea.), e.g. *café*, 'coffee'. These words are accounted for by making the final vowel part of the stem, i.e. not a 'thematic' vowel.
3. Proparoxytonic words ending in a consonant: only 3 words: *régimen*, 'regime', *espécimen*, 'specimen', and *ínterin*, 'meanwhile' (archaic).[11] These words are accounted for by labeling the last two vowels as unfit to receive stress.

Hooper & Terrell's account seems similar in spirit to later metrical ones which use the notion of extrametricality (cf. Harris 1983), and does a reasonable job of capturing the *surface* generalizations about the stress of Spanish non-verb forms.

Lest it be thought that all recent accounts of Spanish stress are rule oriented, I should mention a paper by Jeff Elman (1979) in which he argues against Harris' (1969) phonological penultimate stress rule on the basis of a variety of 'external' evidence: double plurals, loan words, psycholinguistic tests, and historical and dialectal change, and proposes that "in Spanish there are classes of words with distinctive stress patterns. Class membership might be determined on the basis of several factors; phonological shape is certainly not irrelevant, but may be secondary to morphological structure or semantic analysis" (p. 12). In other words, "Spanish noun and adjective stress is not governed by a single rule. It is the result of a complex process by which a word is assigned to a stress class" (p. 15). Thus, for him stress is stored in the words in the lexicon and word class is motivated by a variety of factors.

3. Testing the psychological reality of Spanish stress rules: the hypotheses

The generativists' claim (if we are to believe their stated concern with psychological reality) is thus that Spanish speakers make a (phonological or morphological) generalization about the surface stress patterns of the language and mark the exceptions somehow in the lexicon. It would be desirable not to credulously accept this assumption and to test whether this is in fact the case. The obvious alternative to pattern extraction, abstraction, or "disembodiment", would be that stress is simply stored with the words in the lexicon. Let us refer to this view as the **Patterns in the Lexicon** (PL) view, as opposed to the generativist or **Disembodied Rule** (DR) view. This possibility was already envisioned for instance by Chomsky when he said: "Phonology, as distinct from syntax, is a system that is essentially finite in scope. It would be possible, in principle, for the mapping from surface structure to phonetic representation to be simply memorized, case by case" (Chomsky 1967, pp. 126-7; cited in Derwing 1973, p. 123). The implications of the resolution of this controversy go beyond the issue of Spanish stress, for if a phonological phenomenon as

regular (but not automatic) as Spanish stress can be shown to be simply stored and not abstracted in the form of a rule by speakers this would have serious consequences for our view of phonology and its psychological implementation.

One particular area in which these two models, the DR model and the PL model, would seem to be making differing and testable predictions has to do with how speakers would react to new words presented to them in such a way that they would have to choose the appropriate stress pattern. Taking advantage of the fact that in capitalized script in Spanish stress is conventionally left unmarked, we can express the different predictions as follows:

HYPOTHESES/PREDICTIONS:
If speakers encounter a made-up word written in capital letters which they have never seen before, then:
* **DR model**: Since they have made a generalization of the type expressed in the stress rules, then they will use this generalization and assign stress accordingly.
* **PL model**: They will access the lexicon itself to assign stress to the new word using some sort of analogical process, and, when looking for a suitable pattern, they _may_ be influenced for instance by
  i. the specific phonological shape of the word they are to assign stress to (rather than the general shape the rules say is relevant), or by
  ii. other non-phonological information about that word (that, again, the rule does not predict is relevant).

Major patterns		Minor patterns	
-a:	33.053%	-t,-d:	0.919%
-o:	26.390%	-i,-j:	1.670%
-e:	9.684%[1]	-V:	1.630%
-s:	6.593%[2]	Other:	1.159%
-r:	11.492%[3]	Total:	5.378%
-l:	2.554%		
-n:	4.856%		
Total:	94.622%		

Notes
1. Many are adverbs in _#mente_.
2. Not counting plurals (in -(e)s) and finite verb forms.
3. 71% of these are infinitival forms of verbs (-{a,e,i}r).

Table 1 Segments found in Spanish word endings. Not included: verb forms except infinitives (all in -r) and plurals (in [...V]-s, [...C]-es). Source: Faitelson-Weiser (DIASLE), p. 177.

A specific area where the two models are likely to make different predictions is one where phonological facts not present in the rule analyses are associated with a different stress pattern. I have in mind here, for instance, the fact that although paroxytonic words ending in a consonant (a limited number of consonants are available word finally in Spanish, see Table 1) are in general exceptional,[12] this is not true of words ending -en. As can be seen in Table 2, although the vast majority of words ending in -n are oxytonic, words in -en are around 62% paroxytonic, or around 50% if we discount some quite rare words. By words 'familiar to the author' I mean words that I thought, using a subjective criterion, were not exceedingly learned or archaic and thus were likely to be at least vaguely familiar to an educated speaker. These facts lend themselves to formulating the following corollaries to the hypotheses or predictions made above:

COROLLARIES (1):
If speakers encounter a new, written, made-up word ending in -en:
* **DR model**: they will follow the rule/generalization; thus, since the word ends in -n (a consonant), stress it on the final syllable.
* **PL model**: they will examine their mental lexicon for a suitable model/pattern there; thus they _may_ be influenced by the fact that penultimate stress is a distinct possibility for words in -en.

	Total in dictionary		Familiar to the author	
	Penultimate	Final	Penultimate	Final
-an	15	338	1[1]	54[2]
-in	1	479	1[3]	62
-on	22	4,463	11[4]	n/c[5]
-un	0	25	0	6
-en	76	156	34[6]	21[7]

Notes (n/c: not computed; a high percentage)

1. *ránglan* (also *ranglán*, both rare). In DIASLE there are 3 additional borrowings from English: (?)*púlman*, (?)*bárman*, and *eslógan*.
2. In addition there are 4 monosyllabic words.
3. *esmóquin* ('smoking jacket'). DIASLE also has (?)*mítin* (~*mitín*), < *Eng.* meeting.
4. 5 somewhat rare paroxytonic words: *náilon, cánon, pláncton, néwton, cláxon*; and 6 quite rare proparoxytonic words: *épsilon, ómicron, hipérbaton, asíndeton, polisíndeton, rémington*. (DIASLE also has *Nélson* and (?)*bádmington*.)
5. 2,168 in *-ción* (mostly common deverbal nouns).
6. 32 are paroxytonic: of which 22 are **common** words: *órden, contraórden, desórden, imágen, orígen, aborígen, márgen, vírgen, álguien, pólen, certámen, exámen, sémen, hímen, crímen, dólmen, abdómen, Cármen, gérmen, volúmen, resúmen, jóven*; and 10 words are more *rare: ligámen, vejámen, pelámen, velámen, dictámen, gravámen, albúmen, cacúmen, glúten, líquen*. 2 words are proparoxytonic: *espécimen, régimen*.
7. *Jaén, almacén, arcén, Almadén, Edén, andén, desdén, rehén, también* (Adv.), *recién* (Adv.), (?)*parisién* (~*parisino/a*), *Jerusalén, Matusalén, Belén, terraplén, amén, santiamén, harén, retén, sartén, sostén, vaivén*. (Also *arcén*).

Table 2 Number of words in *-Vn* and *-V̄n*. Source: Bosque & Pérez Fernández.

Another 'semi-sub-regularity' of Spanish stress concerns words ending in *-ico* and *-ica* (*-ic{o,a}*). As can be seen in Table 3, the vast majority of words in *-ic{o,a}* have proparoxytonic (antepenultimate) stress, for they are historically forms containing the adjective forming suffix *-ic-* (which is stress-retracting in the rule accounts). However, a count of these words from Bosque & Pérez Fernández's reverse dictionary reveals that approximately 43% of the proparoxytonic words ending in *-ica* are either not analyzable at all (even to an educated native speaker) as consisting of a stem plus *-ic{o,a}*, or only partially analyzable (e.g. *síntesis ~ sintética*), and of these, about 76% are adjectives (e.g. *fanática, 'f.* fanatic'), about 19% are both nouns and adjectives (e.g. *química, 'a.* chemical', *'n.* chemistry'), and only about 5% could be only nouns (e.g. *república,* 'republic'). Also, there are a few paroxytonic words in *-ic{o,a}*. The question is: What is the generalization that native speakers extract from this lot of words? Could it be that speakers make a generalization having to do with the fact that most of these words are adjectives? Our two models could be said to make the following predictions about the relevance of word class information:

<u>COROLLARIES (2):</u>

If speakers encounter a new, written, made-up word ending in *-ic{o,a}* (not the suffix):

* **DR model**: they will follow the general rule; thus since there is no evidence for a stress retracting suffix and word class (N vs. A) information is not relevant for stress assignment, they will stress the word on the penultimate syllable.
* **PL model**: they will consult their mental lexicon in search for a similar pattern, and so word class information (N vs. A) may be one of the relevant factors. Thus they

	Total	Familiar	Compositional	Noncompositional
$-\acute{V}C_0$ica	1,743	663	358[1]	275[2]
-íca	53	12[3]	n/a	n/a
$-\acute{V}C_0$ico	1,837	n/c	n/c	n/c
-íco	68	7[4]	n/a	n/a
-icó	2[5]	1	n/a	n/a

Notes (n/a: not applicable; n/c: not computed)
1. 51 words in *-graf{o,ía}+ica*, e.g. *telegráfica* (cf. *telégrafo, telegrafía*) and 57 in *-log{o,ía}+ico*, e.g. *teológico* (cf. *teólogo, teología*).
2. 210 adjectives, e.g. *láctica, semítica*; 14 nouns, e.g. *república, encíclica, basílica, crónica, Verónica, fábrica, América, música, plática*; and 51 noun/adjective, e.g. *química, técnica, clínica, lírica, táctica, política, ...*
3. 6 nouns: *chíca, caníca, Dominíca, picapíca, Federíca, barríca*; 1 adjective: *ríca*; 5 noun/adjective (colloquial & compositional): *quejíca, roñíca, maríca, lloríca, acusíca*. In addition the dictionary has 15 dialectal diminutives (in *-íc-* instead of *-ít-*; all nouns and rare): *herbecíca, tardecíca, ternecíca, fontecíca, portecíca, cieguecíca, mañaníca, vainíca, boníca, risíca, setíca, chiquirritíca, cinturíca, borríca, botíca.*
4. (4 are bisyllabic!) *hocíco, zorcíco, chíco, míco, abaníco, píco, ríco*. An additional 13 are dialectal diminutives (in *-íc-* rather than *-ít-*): *ternecíco, cornecíco, corpecíco, ventecíco, cieguecíco, callandíco, caballíco, boníco, asníco, aceríco, majaderíco, Períco.*
5. (?)*calicó, Jericó*.

Table 3 Number of words in *-ico, -ica*. Source: Bosque & Pérez Fernández.

may be more likely to stress adjectives in *-ic{o,a}* on the antepenultimate syllable than they are nouns in *-ic{o,a}*, since most words in *-ic{o,a}* are adjectives.

4. The experiment

In order to test these different predictions about how native speakers would react to made-up words ending in *-n* and in *-ic{o,a}* they had not encountered before, a list of such words in context was designed. The words in *-n* that were used (6 pairs) can be seen in Table 4. The test words (1-6) end in *-en*, and the control words (7-12) end in a different vowel plus *n* (-{a,i,o,u}n). The results can be seen next to the words: 1 means final stress, 2 penultimate, and 3 antepenultimate. The 6 pairs of words in *-ic{o,a}* and their results can be seen in Table 5. Half of them had to be interpreted as adjectives in context (13-18, left column) and the other half as nouns (19-24, right column). The 24 sentences were randomized and then 12 sentences with real but rare Spanish words were interspersed every two words. These 12 'foil' words, with their results, can be seen in Table 6. Thirty six native speakers of Spanish[13] read the 36 sentences (see Appendix), which were presented in bound booklet form, with one sentence per page. The numbers in the columns labeled *S#* in Tables 4-6 stand for the order the word came in the test.

The results of the two tests can be seen comparatively in Figures 1 and 2, as well as in Tables 4 and 5. In short, he outcome of the *-en* test supports the PL model's prediction, with speakers stressing *-en* words with penultimate stress 43.5% of the time and with final stress 55.6% of the time (one word, CORUMEN, was given antepenultimate stress by two subjects), whereas the other words in *-n* were given final stress 96.8% of the time ($\chi^2 = 8.289$; significant, $p > 0.01$). As for the *-ic{o,a}* test, although the results point in the direction predicted by the PL model, the difference is not statistically significant ($\chi^2 = 1.826$; $p > 0.2$): Adjectives: Antepenultimate--83.3%, Penultimate--16.7%; Nouns:

S#	-en	2	%	1	%		S#	other -Vn	2	%	1	%
1 (11)	BESOREN	8	22.2	28	77.8		7 (3)	SEBORAN	1	2.8	35	97.2
2 (23)	CORUMEN	23	63.9	11	30.6		8 (8)	PORUBON	0	0.0	36	100.0
3 (33)	PETABEN	4	11.1	32	88.9		9 (18)	PETAMIN	0	0.0	36	100.0
4 (17)	FADEN	21	58.3	15	41.7		10 (35)	TEDON	0	0.0	36	100.0
5 (36)	GORQUEN	12	33.3	24	66.7		11 (5)	SORQUIN	1	2.8	35	97.2
6 (14)	MERASEN	26	72.2	10	27.8		12 (20)	PERASUN	5	13.9	31	86.1
	Total	94	43.5	120	55.6			Total	7	3.2	209	96.8

Chi-square: **8.289**, p > 0.01 (significant)

Table 4 Made-up words in *-n* and their results. 1-6: *-en*; 7-12: *-{a,o,i,u}n*. S#: order of sentence containing the word in the test; CÓRUMEN: 2 ea. or 5.5%.

S#	Adjectives	3	%	2	%		S#	Nouns	3	%	2	%
13 (29)	DRONICO	34	94.4	2	5.6		19 (15)	PROMICO	17	47.2	19	52.8
14 (2)	SEPAJICA	25	69.4	11	30.6		20 (24)	NERAJICA	15	41.7	21	58.3
15 (6)	MELICO	32	88.9	4	11.1		21 (30)	LEGICO	31	86.1	5	13.9
16 (21)	DARSICA	34	94.4	2	5.6		22 (26)	LARSICO	32	88.9	4	11.1
17 (32)	FELULICO	23	63.9	13	36.1		23 (9)	MESULICO	14	38.9	22	61.1
18 (12)	CLAPERICO	32	88.9	4	11.1		24 (27)	PRATERICO	28	77.8	8	22.2
	Total	180	83.3	36	16.7			Total	137	63.4	79	36.6

Chi-square: **1.826**, p > 0.2 (not significant)

Table 5 Made-up words in *-ic{o,a}* and results: 13-18: Adjectives, 19-24: Nouns.

S#		3	2	1		S#		3	2	1
25 (1)	REOMETRO	24	12	0		31 (19)	PUSILANIME	36	0	0
26 (4)	SABICHOSO	0	36	0		32 (22)	ESCARAPELA	0	36	0
27 (7)	MORCON	0	1	35		33 (25)	MENJUNJE	0	36	0
28 (10)	MORDAGUERA	0	36	0		34 (28)	ARANDELA	2	34	0
29 (13)	PARLERO	0	36	0		35 (31)	ZAFANTE	0	36	0
30 (16)	PERVINCA	0	36	0		36 (34)	SUBILLA	2	34	0

Table 6 Real but rare 'foil' words used for padding in the test in positions 1, 4, 7,...

Antepenultimate--63.4%, Penultimate--36.6%.

5. Discussion

5.1. Analysis of the results. The result of the *-en* test shows that speakers treat words in *-en* differently from words in *-an*, *-in*, *-on*, and *-un*, *reflecting the actual patterns in the lexicon quite closely.* This result is incompatible with the hypothesis that speakers make an abstract generalization about Spanish stress, along the lines of Harris' or Hooper & Terrell's rule analyses, since these minor 'sub-patterns' do not in any way figure in the formulation of these rules. Furthermore, it is not easy to see how this minor pattern could be incorporated into an abstract generalization or rule that doesn't make direct reference to the lexicon, since the sub-pattern (paroxytonic words in *-en*) is **not categorical, but variable:**
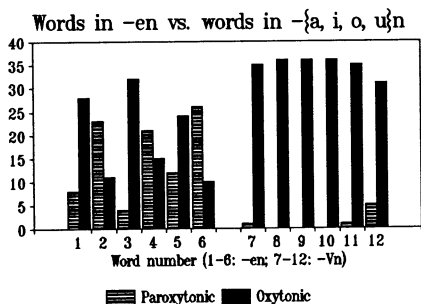
Figure 1 Test 1: *-en* words (1-6) vs. *-{a,e,o,u}n* words (7-12).
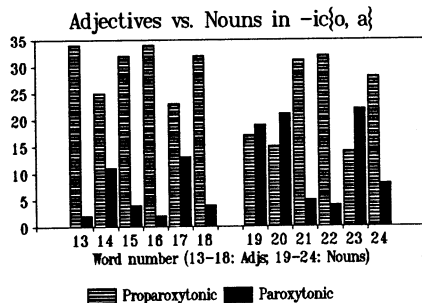
Figure 2 Test 2: *-ic{o,a}* adjectives (13-18) vs. *-ic{o,a}* nouns (19-24).

only about half of the words in -en *in the lexicon do not follow the oxytonic pattern for words in* -n. That is, there is no generalization one can make about the stress of words in *-en* (be it in terms of extrametricality or some other equivalent formalism), like the one possible for most words in *-n* (i.e. final stress).

The result of the second test (the *-ic{o,a}* test) fails to support the prediction that speakers treat nouns in *-ic{o,a}* differently from adjectives in *-ic{o,a}*. That is, although there is a difference which points in that direction, it is not large enough to be statistically significant. However, one thing the results reflect is that speakers treat words in *-ic{o,a}* very differently from other words ending in a vowel, i.e. they stress them overwhelmingly on the antepenultimate syllable instead of the penultimate one, in spite of the fact that these 'words' cannot possibly be analyzed as consisting of a stem plus the suffix *-ic-* plus a theme vowel. This is exactly what the PL model would have predicted. Although this was not part of the test, and control words in, say, *-ac{o,a}* and *-ec{o,a}* were not provided, it is very likely, and a question for further tests, that if they had been and the following corollaries proposed, the results would have been as predicted by the PL model:

COROLLARIES (3):
If speakers encounter a new, written, made-up word ending in *-ic{o,a}*:
* **DR model**: they will access the general rule/generalization and,
  i. If there is evidence for a suffix *-ic-* they will stress the word in the antepenultimate syllable.
  ii. If there is no evidence for a suffix *-ic-* they will follow the general pattern and since it ends in a vowel, they will stress the word on the penultimate syllable.
* **PL model**: they will consult the lexicon directly for a matching pattern, and, since a majority of unanalyzable words ending in *-ic{o,a}* have antepenultimate stress, they may, by analogy, assign this word antepenultimate stress.

5.2. Questions about the experiment. The whole idea of resolving theoretical questions by means of 'external' evidence in general, and experimentation in particular, is not a new one, but it is one that is not thought of very highly among generative phonologists, or at least not enough to put it into practice. I will not address the issue of the validity of experiments to settle theoretical questions, which I take for granted, and will concentrate on problems, or potential problems, with the experiment in order to better understand the variables involved and improving the tests for possible future runs.

One possible reaction to this test is to say that it didn't actually tap the subjects' 'implicit' (rule) knowledge of the patterns in the language and that they merely used some form of analogy, but not because they lack the rules. Although I address related aspects

of this question in the next section, I should say that I did expect this test to tap whatever is there to be tapped about speakers' knowledge of the stress patterns of Spanish. I do not see any reason, other than possible flaws with the experiment, why, if speakers have made an abstract generalization similar to the proposed stress rules for Spanish they would not use such knowledge in the context of this test.

Another issue is that, even if subjects did consult their mental lexicons when called to stress new words, there was no guarantee that they would look at certain features of segments (e.g. -e- vs. -o-), and if they did, of how many segments (e.g. -en vs. -ten). It turned out, however, that there was a rather good correlation between percentage of words ending in -$V_i n$ in the lexicon and the percentage of words in -$V_i n$ stressed like those words in the test. The possibility that a larger degree of similarity (more segments) would be relevant, say one more segment, such as the C_i in -$C_i V_i n$, would be compatible with the fact that *CORUMEN* (the test word in sentence 23) had a very high percentage of penultimate stress, since many of the real words in -en with penultimate stress also end in -men. This logic, however, doesn't help with the other cases.

A related problem is that there was no way of knowing beforehand whether the number of words having a certain subpattern would be a good indicator of the functional weight of such a pattern since we didn't know what the cognitive factors involved were. Some words could have had a higher functional weight due to their being more common, and this could have changed the results greatly. For instance, the fact that many of the oxytonic words in -*én* are proper nouns, unlike the words in the test, could have turned out to play a role. As far as I can see, however, no such factors turned out to be relevant.

An interesting fact that I can't account for is that not all words in a particular subgroup behaved the same way in the test. For instance, the most noticeable discrepancy in the -*ic{o,a}* test is that half of the nouns have very low scores (between 4 and 8%) for penultimate stress, whereas the other half have scores of around 20%. In fact, if we ignore the nouns with very low scores and their corresponding adjectives, the difference between nouns and adjectives in -*ic{o,a}* is statistically significant. Also, for the words in -*en* there are two words that were given predominantly final stress, whereas the other 4 were stressed on the penultimate syllable between 33 and 72% of the time. The reasons for this asymmetry are also not clear.

Finally I should mention that groups of words ending in a consonant other than -*n* in Spanish lend themselves to similar tests. For instance, although nouns and adjectives in -*r* are predominantly oxytonic, those in -*er* seem to be about 50% paroxytonic. Also, the percentage of paroxytonic words in -*il* is much higher than the percentage of words in a different vowel plus -*l*.

5.3. Rules vs. non-wasteful listing. As we have seen, the results of the -*en* test support the hypothesis that Spanish stress patterns are stored in the lexical entries themselves and that speakers do not have lexicon-independent abstracted generalizations about such stress patterns. The results of the -*ic{o,a}* test do not support the hypothesis that grammatical (word class) information, i.e. adjective vs. noun, is relevant for assigning stress to a new word in Spanish, although they point in that direction. Indirectly this test also suggests, however, as we saw in section 5, the same support of the hypothesis as the -*en* test. These results suggest that the surface patterns of a language, and in particular the abstract characterizations of those patterns made by linguists, are not necessarily isomorphic with the cognitive or mental structures that produce those patterns.

These results are bound to cause either concern, or, alternatively, denial, among linguists who are most deeply influenced by the **list vs. rule** metaphor of cognition which says something like: 'In description as well as in cognition, something is listed if it is totally

arbitrary and unpredictable; otherwise, it is predicted/abstracted by a rule.' It should be clear by now that such a model of human cognition is extremely naive and that the mind does not work in such a limited all-or-none fashion. This is not the first time that the assumption that all regularities in language are a sign of an abstract rule/generalization reached (implicitly) by the speaker has been challenged, and, as some have argued, in particular proponents of what Derwing & Skousen (1989) call 'analogical' models, just because something is listed it doesn't mean that the (lexical) redundancies are not taken advantage of by the cognitive system in storing, retrieving, etc., without having to resort to abstractions of any sort.

One area of phonology in which regularities have less often been attributed to abstract rule knowledge because of the difficulties involved is that of phonotactics or 'morpheme structure constraints'. Such constraints typically prove difficult to write as rules and thus to extract from the lexicon, although where possible linguists haven't hesitated to do so, as with Spanish *₍ₛ/sC and 'epenthetic' e- (cf. Harris 1969). John Ohala & Manjari Ohala (1986), however, carried out experiments which suggest that speakers consult the lexicon directly and not disembodied rules when making judgements about morpheme structure constraints.

Another study which supports lexical storage of redundancies is the one by David Rumelhart & James McClelland (1987), in which they look at the acquisition and storage of past tense forms of English verbs. In their paper they argue for a realistic model of cognition, grounded on the functioning of the brain, in which the regularities and redundancies of the stored forms are not stored there 'wastefully', but rather the patterns and subpatterns reinforce the storage networks and assist in the organization, storage, retrieval, access, etc. of the items. This model of the cognitive machine seems to me to be more realistic than the rule vs. list model, the latter being nothing but the linguist's useful tool, a simple initial hypothesis for "characterizing, at least approximately, what [speakers] will and will not say" (p. 195). Rumelhart and McClelland argue that rules are rather inadequate to account for human learning and actual use of language, which is much more probabilistic in nature, and which can tolerate enormous amounts of variability and idiosyncracy meshed with subregularities in the system.

Once we have a sophisticated storage model of this sort it is easy to see that creative or novel uses of a pattern may be accounted for by a sophisticated version of what has for a long time been known as analogy, at least until the notion was disparaged by generativists in favor of rules to account for novel uses of existing patterns. But analogy is not the caricature that generativists have made it to be (cf. Derwing & Skousen 1989). For instance Gustaf Stern, well before the era of modern linguistics, foresaw that analogy was more than the creation of a form on the basis of a single form (proportional or four-part analogy) but a rather more complex phenomenon: "A new formation is not necessarily made on the basis of an individual pattern. It is more likely to be guided by an abstract scheme..." (Stern 1965[1931]:211).

5.4. An alternative intermediate solution. Joan Bybee & Dan Slobin (1982) have argued for an alternative to storage and to 'rule regularity', namely 'schema regularity', as reflected in the morphophonemic alternations of English strong verbs, e.g. *keep~kept, leap~leapt, sleep~slept.* According to them, speakers do not treat such sub-regularities as rules, nor do they merely list them. Although they are "rote-learned and stored in the lexicon,"
this does not prevent speakers from formulating generalizations about these forms. These generalizations are not in the form of rules that derive one thing from another by changing features. Thus we will not call them rules, but will rather refer to them as SCHEMAS. A SCHEMA is a statement that describes the phonological properties of a morphological class (in this case past tense). It does not relate a base form to

a derived one, as a rule does, but describes only one class of forms (the product class, in terms used by Zager 1980). It is not a constraint which rigidly specifies what can and cannot occur, but is rather a much looser type of correlation, used in organizing and accessing the lexicon. (pp. 266-7)

One schema that they claim to have evidence for can be stated as either (4a) or (4b):

(4) a. A past-tense form may have the vowel [uw]

  b. ...uw]$_{VERB, PAST}$

as in *drew, blew, fly*. Note "that the base forms of these verbs (*draw, blow, fly*) do not all have the same vowel, so the class cannot be defined in terms of the [phonological] input, nor can a rule be stated that changes a single base vowel into the vowel [uw]. Rather, what these verbs share is the single vowel in their past forms" (p. 267).

If we accept the existence of such a mental entity as the schema, i.e. an abstract (disembodied) statement of a regularity different than a rule, it might be argued that Spanish speakers have one such schema for words ending in *-en*, for instance, which overrides the general phonological rule and which says something like: 'Words ending in *-en* often have penultimate stress.' (Or, alternatively: 'The *-n* of words in *-en* is often extrametrical'.) Although to me this is more plausible than a rule account of such a semi-sub-regularity, I see no reason, on the basis of this test, to posit any such intermediate, abstract representation, rather than assuming that the lexicon is consulted directly.

This doesn't mean that schemas may not turn out to be relevant. I suspect, as more and more linguists do, that schemas, and not rules, are bound to be the best conceptual tool to capture all kinds of mental linguistic abstractions, from phonology, to syntax, to semantics. Schemas have the advantage of being much more flexible--unconstrained some might say--than rules, being rather like templates which characterize all kinds of patterns found in language, typically, but not exclusively, grammatical constructions, with all the relevant formal, semantic and pragmatic information associated with them. (cf. Fillmore 1988, Fillmore, Kay & O'Connor 1988). Thus, it could be even argued, schemas are not an alternative to rules, but rather, rules are a special and more general type of schema.

6. Excursus: Parallel phenomena in syntax

Generative linguistics is based on the proposition that there are very general underlying principles, innate or otherwise universal, which can be violated only at a cost. Two classical examples of this are X-bar principles and grammatical relations. I think that our results in the area of phonology add force to those who argue for alternative ways of conceiving of these topics. I have in mind here in particular the notion of grammatical construction in Construction Grammar (cf. e.g. Fillmore, Kay & O'Connor 1988) and the notion of grammatical relations in Role and Reference Grammar (cf. Foley & Van Valin 1984).

According to X-bar syntax, there are general rules for the expansion of different types of phrases and sentences, or at least a few such types of expansion crosslinguistically. When a language is not consistent in some way, a special statement is made to this effect (notice the similarity to the notion of phonological rule). In Construction Grammar on the other hand the basic units of grammar are constructions, which go from the most general and underspecified to the very specific, which may be even partially filled in with lexical items, as well as with all the semantic and pragmatic information associated with the constructions. It is not that all constructions in a language have the same status, but they all do belong to the same type of objects. And, just as with the stored stress patterns, the inter-construction redundancies and regularities are not 'wastefully listed', but rather they reinforce each other while stored as separate patterns or schemas. Once this possibility is envisioned, all the appeal of the X-bar model disappears and can be seen for what it is,

namely, an interesting statement **about surface patterns**, not a reflection of internal mental representations.

Another possible parallel has to do with grammatical relations. In Role and Reference Grammar grammatical relations (actually the **pivot** grammatical relation, since there are no others) are not basic, universal categories that remain invariable from language to language or even from construction to construction within a language, which can be 'violated' in a language or in a construction at a cost. Rather grammatical relations reflect construction-specific neutralizations of semantic/functional distinctions for the purposes of syntax. In some languages, such as English, grammaticalized reflexes of this category are remarkably constant throughout most construction types, which makes it reasonable to call them *the* subject of English, but in other languages this is much less so. Although having to state what the pivot is for each construction may seem highly redundant, especially when it is typically highly predictable within a language and even crosslinguistically, this analysis is consistent with the claim that although redundancies may be listed in different places the cognitive system may still recognize them and make use of them.

7. Conclusion

The results of this experiment suggest that speakers of Spanish do not make absolute, abstract generalizations about stress patterns, which are then overridden by lexical specifications, for they do not seem to use any such generalization when encountering new words. Rather, the results suggest that speakers look at the lexicon directly for a suitable pattern, concentrating on the last few segments of a word, since the results mirror the patterns in the lexicon quite closely in their variable behavior. This would imply that, at least non-automatic phonological features of this kind perhaps should be assumed to reside in the lexicon and not in abstract rules disembodied from the lexicon. In other words, the linguist's abstract rule characterizations of surface linguistic patterns are not necessarily the best model for the internal cognitive mechanisms that produce those patterns.

APPENDIX

1. NECESITO UN NUEVO REOMETRO PARA MI AUTOMOVIL.
2. LA ESCENA MAS SEPAJICA DE ESA PELICULA ESTA AL FINAL.
3. DESCUBRIERON UN SEBORAN PREHISTORICO DENTRO DE UNA CUEVA.
4. MI TIO LUIS ERA UN CABALLERO BIEN SABICHOSO.
5. EL SORQUIN ES UN INSTRUMENTO QUE USABAN LOS CARPINTEROS.
6. EXPULSARON AL TRABAJADOR MAS MELICO DE LA FABRICA.
7. A LA LAVADORA LA LLAMAN MORCON EN PANAMA.
8. ANUNCIARON LA ENTRADA DE UN PORUBON EN LA ATMOSFERA.
9. ESTA MAÑANA ME TOME UN MESULICO PARA DESAYUNAR.
10. LA PROFESORA DE QUIMICA PILLO UNA MORDAGUERA MUY FUERTE.
11. EL LIQUIDO QUE SALE DE ESA GLANDULA SE LLAMA BESOREN.
12. SE MOSTRABA MUY CLAPERICO EL EMPLEADO AQUEL DIA.
13. VICTOR ES DEMASIADO PARLERO PARA MERECER ESE TRABAJO.
14. NO DEMOSTRARON QUE EL VIRUS MERASEN FUERA TRANSMISIBLE.
15. EXISTE UN PAJARO TROPICAL QUE SE LLAMA PROMICO.
16. UN ZAPATO PERVINCA UTILIZA BOTONES EN VEZ DE CORDONES.
17. ESTA MODA DE CALZADO ES DEMASIADO FADEN PARA MI GUSTO.
18. DIJERON QUE EL QUE TUVIERA UN PETAMIN QUE SE FUERA.
19. NO ENCONTRE LA PALABRA PUSILANIME EN MI DICCIONARIO.
20. EL PATRON ME MANDO QUE TRAJERA UN PERASUN METALICO.
21. LA MUJER QUE VINO A VERME ERA MUY DARSICA.
22. CADA VEZ QUE VEO A JUAN ME OFRECE UNA ESCARAPELA.
23. ¿A QUIEN SE LE OCURRIRIA TRAER UN CORUMEN TAN GRANDE?
24. MI HIJO QUIERE UNA NERAJICA PERO NO SE LA VOY A COMPRAR.
25. PRETENDIAN VENDER ESE MENJUNJE POR CINCUENTA PESOS.
26. EL GANADOR DE LA RIFA OBTENDRA UN LARSICO NUEVO.

27. MI TIA ANCIANA VIVE EN LA CALLE PRATERICO, NUMERO 88.
28. NO PUEDEN ARREGLAR EL TELEVISOR PORQUE FALTA UNA ARANDELA.
29. TENGO ENTENDIDO QUE JUAN ES UN HOMBRE MUY DRONICO.
30. MI AMIGA PREFIERE UN LEGICO QUE UN APARATO DE MUSICA.
31. EL CONTADOR INDICA QUE LA PROBETA ES DEMASIADO ZAFANTE.
32. ¿POR QUE SERA TAN FELULICO ESE CAMARERO?
33. ME DIJO EL DOCTOR QUE ME TOMARA UN PETABEN CADA NOCHE.
34. INVENTARON UNA SUBILLA QUE ES MUY EFICAZ CONTRA LA POLILLA.
35. PARA ABRIR LA CAJA FUERTE ERA NECESARIO UN TEDON.
36. AUN NO ME HAN MANDADO EL GORQUEN QUE PEDI HACE UN MES.

NOTES

*. I would like to thank John Ohala for the stimulating class on Methods of Phonological Analysis in the Spring of 1986 which led to the pilot study of the experiment reported on here (Aske 1987). For the present experiment I have benefited from comments from Mariscela Amador, Larry Hyman, Paul Kay, John Ohala, Dan Slobin, and María José Solé Sabater, advice which perhaps I should have followed more often than I did.

1. Early structuralists in the US, such as Boas and Bloomfield, as well as some European counterparts, and, of course, the legendary Indian grammarian Panini, had made use of rules and rule-like terminology in their linguistic descriptions.

2. cf. Chomsky's undergraduate thesis (1949) and Halle's (1959) *The Sound Pattern of Russian*.

3. It has been observed, however, that it is precisely the most basic, regular, and redundant phenomena that often provide the most central clues for identifying a phoneme in context, e.g. aspiration of voiceless obstruents and preceding vowel lengthening for final voiced obstruents in English.

4. Phonological features in Generative Phonology enjoy the same virtues and vices of rules. Features can be a great tool for capturing generalizations about what different speech sounds/segments/phonemes have in common, and for writing statements about phenomena associated with sounds sharing something in common. However, when they are turned into theoretical primes in terms of which phonological segments are defined, then probably a generalization is being missed about what speech and actual phonological systems are really like (cf. Ohala 1985).

5. One problem that I see with the weak position is that if rules are not in any way involved in production, then what are the psychological analogues of phonological derivations and the phonological cycle? In Generative phonology, unlike in Generative syntax, these mechanisms do not seem to have been disposed of or curtailed in any way.

6. This view fits well with Chomsky's content view of inborn linguistic structures, but is logically independent.

7. Harris (1975) is a reworking of Harris (1969) in which he stops using the tense/lax distinction, and shifts to a more honestly *ad hoc* feature [X]. Harris (1983) has an analysis of stress framed within the new theory of metrical phonology, which relies heavily on the notion of extrametricality and the existence of class vowels, or class markers (the final vowels of proparoxytones), which are marked in the lexicon (by convention) as extrametrical. Final consonants of proparoxytones which end in a consonant are also extrametrical, as is the penultimate vowel of proparoxytones. As far as I can see this extrametrical account is nothing but a (slightly modified) notational variant of Hooper & Terrell's morphological account, with even less convincing explanatory power.

8. Another type of 'rule' in NGP is the **via-rule**, a lexical rule proposed by Vennemann **1971**, which accounts for "putative alternations of the type *leche* and *lactar*, *noche* and *nocturno*" in Spanish (Hooper 1976, p. 17) (The rule, to the extent that speakers are aware of this correspondence, would be represented in the following way: kt ↔ č).

9. Hooper & Terrell claim that verb stress has a morphological function, i.e. "it is directly correlated with tense" (p. 79), and interesting though not totally convincing claim.
10. As they notice, "[t]he status of /s/ appears uncertain because /s/ marks the plural and pluralization has no effect upon the position of stress in nouns and adjectives." (p. 67) Furthermore, words that end in a consonant, the majority of which have final stress, in the plural have penultimate stress, since an 'epenthetical' /e/ is added to the /s/ (i.e. they make their plural in /-es/).
11. For many speakers the stress shifts in the plural when one more syllable is added, e.g. *regímenes*. The same is true for most speakers with *carácter*, which becomes *caractéres*.
12. The dictionary DIASLE contains quite a large number of rare non-native words absent from other inverse dictionaries, which may make minor patterns (e.g. words ending in an accented vowel) more prevalent than they really are in the vocabulary of the vast majority of speakers.
13. 20 speakers were tested by the author in Berkeley (most of them UC Berkeley graduate students and their spouses) (plus 2 who were disqualified because pages were skipped). Origin: Argentina, 8 ea.; Mexico, 6 ea.; Spanish State, 2 ea.; Chile, 1 ea.; Central America, 1 ea.; Peru, 1 ea.; Cuba, 1 ea. The other 16 (plus 2 more disqualified for the same reason) were tested by the author's associates in the Spanish State. Origin: Euskadi, 14 ea.; Catalonia, 1 ea.; Castile, 1 ea.

REFERENCES

Anderson, Stephen R. 1985. *Phonology in the twentieth century*. Chicago: University of Chicago Press.

Aske, Jon. 1987. "Are there stress rules in the minds of Spanish speakers? Tapping the implicit knowledge of stress patterns." Ms., UC Berkeley.

Bosque, Ignacio & Pérez Fernández, Manuel. 1987. *Diccionario inverso de la lengua española*. Madrid: Editorial Gredos.

Bybee, Joan & Slobin, Dan I. 1982. "Rules and schemas in the development and use of the English past tense." In *Language* 58.2, 265-89.

Chomsky, Noam. 1967. "Some general properties of phonological rules." In *Language* 43, 102-18.

Derwing, Bruce L. 1973. *Transformational Grammar as a theory of language acquisition. A study in the empirical, conceptual and methodological foundations of contemporary linguistics* (CSL 10). Cambridge: Cambridge University Press.

Derwing, Bruce L. & Skousen, Royal. 1989. "Real-time morphology: Symbolic rules or analogical networks?" In *BLS* 15, 48-62.

Elman, Jeff. 1979. "Spanish noun and adjective stress assignment: A non-phonological account." In J.P. Lantolf *et al.*, eds., *Colloquium on Spanish and Luso-Brazilian Linguistics* (1976). Washington, DC: Georgetown University Press.

Ewen, Colin J. & Anderson, John M. 1986. *Phonology Yearbook 3*. Cambridge: Cambridge University Press.

Faitelson-Weiser, Silvia. 1987. *DIASLE: Dictionnaire inverse et analyse statistique de la langue espagnole*. Québec: Les Presses de L'Université Laval.

Fillmore, Charles J. 1988. *On grammatical constructions*. Ms. UC Berkeley. Fall 1988 version used for the course Linguistics 120a.

Fillmore, Charles J., Kay, Paul, & O'Connor, Mary Catherine. 1988. "Regularity and idiomaticity in grammatical constructions: The case of *LET ALONE*." In *Language* 64.3, 501-538.

Foley, William A. & Van Valin, Robert D. Jr. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Harris, James W. 1969. *Spanish Phonology.* Cambridge: MIT Press.

Harris, James W. 1975. "Stress assignment rules in Spanish." In William G. Milan *et al.,* eds., *1974 Colloquium on Spanish and Portuguese Linguistics.* Washington, DC: Georgetown University Press.

Harris, James W. 1983. *Syllable structure and stress in Spanish.* Cambridge: MIT Press.

Hooper, Joan B. 1976. *An introduction to Natural Generative Phonology.* New York: Academic Press.

Hooper, Joan B. & Terrell, Tracy. 1976. "Stress assignment in Spanish: A Natural Generative approach." In *Glossa* 10.1, 64-110.

Kiparsky, Paul. 1968. "How abstract is phonology?" Bloomington, Indiana: IULC.

Linell, Per. 1979. *Psychological reality in phonology. A theoretical study* (Cambridge Studies in Linguistics 25). Cambridge: Cambridge University Press.

Ohala, John J. 1985. "Around *flat.*" In V.Fromkin, ed., *Phonetic linguistics: Essays in honor of Peter Ladefoged,* 223-41. Orlando, FL: Academic Press.

Ohala, John J. 1986. "Consumer's guide to evidence in phonology." In C.J. Ewen & J.M. Anderson, eds., *Phonology Yearbook 3,* 3-26.

Ohala, John J. 1987. "Experimental phonology." In *BLS* 13, 207-222.

Ohala, John J. & Jaeger, Jeri J., eds. 1986. *Experimental phonology.* Orlando, FL: Academic Press.

Ohala, John J. & Ohala, Manjari. 1986. "Testing hypotheses regarding the psychological manifestation of morpheme structure constraints." In J.J. Ohala & J.J. Jaeger, eds.

Rumelhart, David E. & McCleland, James L. 1987. "Learning the past tenses of English verbs: Implicit rules or parallel distributed processing?" In Brian MacWhinney, ed., *Mechanisms of language acquisition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stern, Gustaf. 1965 [1931]. *Meaning and change of meaning. With special reference to the English language.* Bloomington: Indiana University Press.

Vennemann, T. 1971. "Natural generative phonology." Paper read at annual meeting of the Linguistic Society of America, St. Louis, Missouri.

Whitley, Stanley. 1976. "Stress in Spanish: Two approaches." In *Lingua* 39, 301-332.

Zager, David. 1980. *A real-time process model of morphological change.* Buffalo, NY: SUNY dissertation.

Underspecification in American Sign Language Phonology

Diane Brentari
University of Chicago

I. Introduction

This paper is concerned with developing a theory of underspecification in American Sign Language (ASL). The purpose of this analysis is to show that underlying representations in ASL are not necessarily 'whole signs'; instead, these underlying forms often consist of incomplete structures that require syllabification or other prosodic principles in order to make them articulable in the language. Until now, what is seen in the surface form of signs has been considered to be in the underlying form as well. I show that not only can a large portion of this phonetic material be eliminated from underlying representations, but it must be eliminated if we are to arrive at the correct and most economical analysis of the following phenomena. My evidence comes from:

1) Compounding: I show that the input to the compounding process is not the surface form of the two lexical items that form the compound, but rather that the compound is formed from two underspecified representations.

2) Distribution of the feature value [+contact]: Both [+contact] and [-contact] have been posited in the underlying representation of previous work on ASL (Liddell and Johnson 1989). If only [+contact] taken as a part of the underlying representation, we can see how its distribution can be predicted, and how [-contact] is assigned.

3) Signs that undergo what has been called metathesis (Liddell and Johnson 1989). One group of signs allows the two places of articulation (POA) to occur in either order-- POA1, POA2 or POA2, POA1.

The theory I will be using is a 'harmonic' approach to phonology (Goldsmith 1990). In this theory, there are three levels of phonological representation which are present in the grammar in a parallel, rather than a derivational fashion. That is to say, the set of principles operating on a given structure at one of those three levels has access to all of the information from the other two levels of the phonology. A given structure will strive to achieve a "best fit" with all three levels. I will limit my description of these 3 levels to details needed for the analysis to follow. These levels, which are all cognitive levels of the phonological component of a grammar, are:

1. Underlying or M(orphophonemic)-Level: This level interfaces with the morphological component of the grammar. It contains the representations composed of underspecified distinctive features which need not be pronounceable according to the phonotactic constraints of a particular language. Many spoken languages do not allow consonants clusters (for example, Yawelmani doesn't allow three consonants (Kenstowicz and Kisseberth 1979, Archangeli 1984)); Yoruba doesn't even allow clusters of two consonants (Pulleyblank 1988). Such languages often insert a vowel in such cases to avoid this situation. This process of epenthesis does not occur at the underlying level, but rather in accord with principles of syllabification and word formation. An analogy can be made between this process mentioned above in spoken languages and ASL, where two distinct places of articulation are allowable at M-Level, even though they do not conform with the phonotactics at the phonetic level: an intervening movement must be inserted.

2. W(ord) Level: This intermediate level of the phonology contains principles of word, syllable, and metrical foot structure, as well as principles which may differentiate between monomorphemic and polymorphemic words in the way they are syllabified.

W-Level phonotactic statements can insert empty slots on the moraic or the skeletal tier in order to conform with the syllable template. In spoken languages like those alluded to above, the syllabic template inserts an epenthetic vowel that often surfaces as the completely unspecified vowel. In ASL, an unspecified movement value, which is phonetically realized most often as an arc movement, is inserted between two contrastive places of articulation. In contrast, movements present in underlying representation are called 'paths'.

3. P(honetic)-Level: This is the component of the phonology that interfaces with the phonetic component of the grammar. Redundant feature values of two types are filled in here: 1) those that arise as a result of trivial underspecification, and 2) those that arise as a result of non-trivial underspecification. (Steriade 1987) Non-trivial underspecification involves segments which lack underlying values for a feature. For example, the distinction between [+aspirated] vs. [-aspirated] stops in English as a result of their position in the syllable and an obstruent stop will acquire this distinction in this way at P-Level. In ASL, the distinction between 'top' and 'bottom', such as [cheek[+top]] vs. [cheek[-top]] in signs such as DEAF and HOME, or [trunk[+top]] and [trunk[-top]] in RELIEF, SATISFIED are acquired as the result of non-trivial underspecification in much the same way as [±aspiration] in English. [1] Trivially specified redundant values can arise as the result of privative underlying features. For example, in languages with a canonical 3-vowel system, vowels that are not [+low] (namely, /u/ and /i/) are filled in with [-low]. Another source of such redundant values is the mechanism responsible for filling in [+voice] for all [+sonorant] segments in English. In ASL, segments that are not marked underlyingly for the feature [contact] or [distal] acquire [-contact] as the result of trivial underspecification.

The underspecification theory assumed here is Restricted Underspecification; that is, a theory that allows all contrastive features of a language to be expressed underlyingly. This is different from Radical Underspecification, which allows only privative features in underlying representation, and only those features that are universally marked. Moreover, I am adopting a mechanism whereby features will be specified as to whether they are privative, equipollent or n-ary underlyingly and may have an unspecified value, as well. In this way, privative features have 2 possible values, equipollent values have 3 values possible, and n-ary features have (n+1)possible values.

An example of how a feature in ASL can be contrastive for some signs, yet not for others, is the [± contra] distinction (an equipollent feature), and it is analogous to [±voice] in English, which is contrastive in obstruents but no where else. In signs that the [± contra] distinction is not contrastive, it is not specified underlyingly. In this analysis, only underspecified, contrastive features will be present underlyingly. Below are examples of signs that are not contrastive for [±contra] distinction in (1a), and examples where it is contrastive in (1b).

1a. *unspecified for [±contra]*	1b. *specified for [±contra]*
FLOWER	DEAF
MEMBER	BEE
NAVY	PITTSBURGH
HONEYMOON	LEATHER
CONGRESS	CHARACTER
SATISFIED	MILITARY

I exploit such redundant contrasts as [± contra] and argue for an underspecified representation in ASL that is more economical than the underlying representations posited by Sandler (1987a) or Liddell and Johnson (1989).

II. Arguments
A. Compounding

In 1978, Supalla and Newport published their ground-breaking article concerning the difference in structure between nouns and verbs in ASL. Many such pairs exist. The verbs in this group form a group of single, punctual actions:

2a.		2b.	
	FLY		AIRPLANE
	SIT		CHAIR
	PUT-ON--HAT		HAT
	CALL		NAME

A significant, but often overlooked idea raised in this article is that neither verbs or nouns are derived from the other, but instead both are derived from a common underlying form. This analysis is driven by that idea. Here, both SIT and CHAIR have the same form underlyingly: (HS=handshape; the lines associating the features to the central morpheme node are not association lines; they refer to the unordered arrangement of the features of the morphological unit in question.)

3.

SIT/CHAIR

[HS[H]]

morpheme

[+contact]   [direction[<--]]
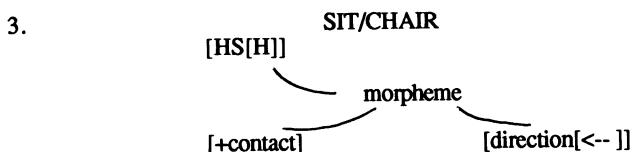
There is no specification for place of articulation because SIT (and CHAIR) are articulated in the unspecified place of articulation--the place called neutral space by Stokoe , et. al. (1965).[2] Because this information is sufficient for syllabification, this structure will syllabify as if there are no other morphophonological structures in play. This is the case in verbs. As Supalla and Newport state, nominalization calls for [repeated] [restrained] manner in articulating the path--here [repetition[x2]]. This results in a bisyllabic structure at W-Level.[3] The syllable structure of SIT and CHAIR are shown in (4).

4a.

SIT

σ

1
2   3

1. [+ contact]

2. [HS[H]]

3. [direction[<--]]

4b.

CHAIR

σ   σ

1
2   3   2   3
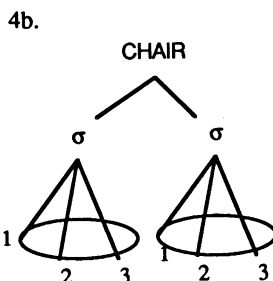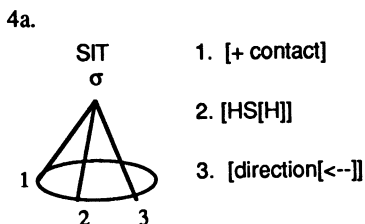
In many compounds, one of the two stems which make up the compound come from such pairs. One such case is the compound often glossed BLACK-NAME=BAD REPUTATION.[4] It is misleading to gloss this sign this way because NAME implies the nominal form, and both Sandler (1987a) and Liddell and Johnson (1989) have posited that the compounding process takes this reduplicated form above, rather than the underlying form, as the input to the compounding process. For Sandler, there is a rule of truncation for reduplicated signs in compounding that operates from the right. This formulation is represented below (Sandler 1987:278).[5]

5.   **BLACK**   **NAME**   **BLACK-NAME**
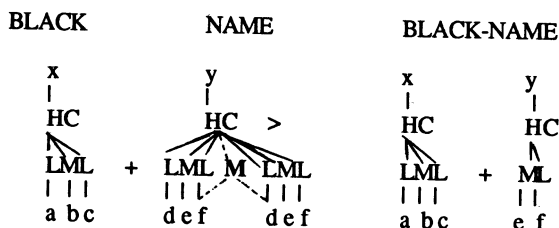
In my analysis, the underlying forms of BLACK and NAME are syllabified directly, yielding a bisyllabic structure, based on the fact that there are two distinctive path features in the respective stems. That is to say, instead of the compounding process [BLACK]word + [NAME]word > [BLACK-NAME]compound, my analysis allows for [BLACK]stem + [NAME]stem > [BLACK-NAME]compound. Both [tracing] and [direction] are path features. The underlying structures are as follows:

6a   [tracing[str]]   [+contact]   6b. [direction[<--]]

BLACK[stem]   NAME[stem]

[POA[forehead]]   [HS[1]]   [HS[H]]   [+contact]
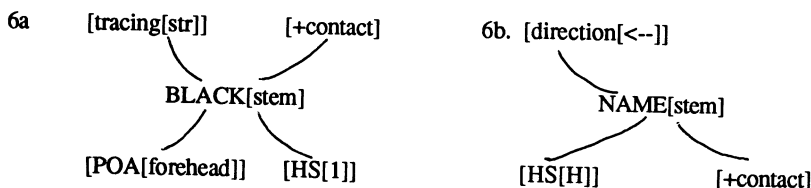
These structures are syllabified at W-Level as follows:

7.

[BLACK[stem]-NAME[stem]compound]]

1. [tracing[str]]   5. [direction[<--]]

2. [+contact]   6. [+contact]

3. [[HS[1]]   7. [HS[H]]
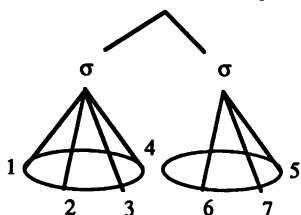
4. [POA[forehead]]

The analysis above suggests a test for predicting the monosyllabic or bisyllabic nature of compounds in ASL. If there are a total of two path features in both stems, the compound will be bisyllabic; if there is just one path feature in both stems the compound will be monosyllabic. Below, some bisyllabic compounds are listed in Group 1; monosyllabic compounds are listed in in Group 2. [6]This evidence suggests that, in the syllabification of compounds, underlying paths are differentiated from other types of movements.[7]

8.   *Group 1*
       SLEEP-SUNRISE ('oversleep')
       BLACK- NAME ('bad reputation')
       SLEEP-SHOES ('slippers')
       THRILL-INFORM ('entertainment')
       NAME-SHINE ('good reputation')

       Group 2
       GOOD-NIGHT ('good night')
       SPEAK-NAME ('mention')
       TOMORROW-MORNING ('next day')
       RED-SLICE ('tomato')
       BLUE-SPOT ('bruise')
       THINK-SELF ('decide yourself')

The contrasting examples in (8) are, on one hand, BLACK-NAME ('bad reputation') and NAME-SHINE ('good reputation'), which have two stems containing two path features, making these compounds bisyllabic, and, on the other hand, SPEAK-NAME ('mention') which contains only one path feature in the two stems and consequently is monosyllabic. It makes no difference whether CALL/NAME occurs as the first or the second element of the compound, it contains the [direction] feature in both cases.[8]

B. Distribution of the Feature Value [+contact]

In this section (B) and in the next section (C) of this analysis, 4 groups of signs will figure prominently: (UR=Underlying representation)

9.

	Group 1	*Group 2*	*Group 3*	*Group 4*
	THINK	GOOD	SHAVE	DEAF
	MY	GOAT	WE$_1$	WE$_2$
	KNOW	KING	FORGET	NAVY
UR	HS	HS	HS	HS
	[POA]	[POA$_1$]<[POA$_2$]	[POA]	[POA]
	[+ contact]	[++contact]	[+ contact]	[++ contact]
			[tracing]	

We will be focusing on Groups 3 and 4 first. Group 3 contains signs that have an underlying path feature--[tracing], as well as [+contact], handshape features, and a POA, while signs in Group 4 have no underlying path, while Group 4 doesn't. Group 4 has [++contact] while Group 3 signs have only [+contact]. These contrasts are seen below in SHAVE (Group 3) and DEAF (Group 4).

10a. SHAVE   10b.   DEAF

[tracing[str]]   [HS[Y]]   [HS[1]]

              morpheme   morpheme

[+contact]   [POA[cheek]]   [++contact]   [POA[cheek]]
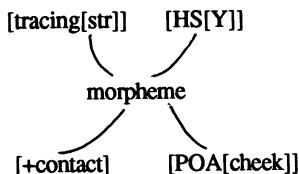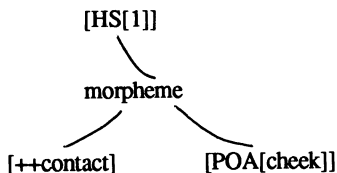
In this analysis, [+contact] is a feature value of the underlying representation, which is defined as follows:[9]

11. [contact]--the feature that indicates that a sign contacts the body or non-dominant hand.[10] The values of this feature are:
   a. [+contact]--segments that contact the body once
   b. [++contact]--segments that contact the body twice
   c. [-contact]--the unspecified value inserted at P-Level.

Handling the feature [+contact] in this analysis hinges on the structural differences underlyingly between Groups 3 and 4. The [contact] value will be associated to the underlying structure. If there is an underlying path feature, [+contact] will associate to that path, otherwise a [+contact] value will associate to the POA. It is noteworthy that, in the signs in Group 3, contact is obligatory during the path, but this contact is not obligatory on the beginning and end points of the path. (The two variants of SHAVE are acceptable.)

Predictably, [+contact] is obligatory at the POA's for signs like DEAF. This explains the difference between Group 3 and Group 4 along this dimension.

Previous work has analyzed [contact] differently. In Liddell and Johnson's system of phonological representation (1985, 1989), both '+' and '-' contact are present in the underlying representation of a single sign. Moreover, for signs such as KNOW, THINK, and MY, both the approaching movement and the contact are present:

12.

$$\begin{array}{ccc} & \text{THINK} & \\ & M & H \\ & 1o\text{-} & \cdot \\ (\text{i. e. .[-contact]}) & \text{proximal} & \text{contact} \ (\text{i. e., [+contact]}) \\ & FH & \\ & \vdots & \vdots \\ & \vdots & \vdots \\ & \vdots & \vdots \end{array}$$
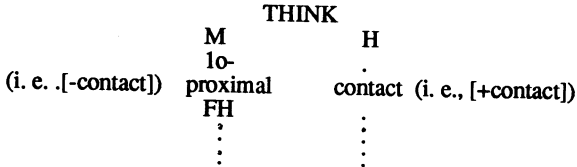
Sandler's (1987a) representation of Group 1 though 4 is identical, and consequently has a difficult time representing the distribution of the [+contact] feature in Groups 3 and 4. This representation on the skeletal tier (here the Location/Movement Tier) makes no distinction between underlying movements --here called paths-- and movements inserted either by the syllable template or the P-Level phonotactics. Even though Sandler recognizes that only [+contact] is present underlyingly, her skeletal template makes predicting its distribution problematic. [Contact] must be specified on some movements but not others. Her representations for DEAF and SHAVE are shown in (13):

13.

DEAF

SHAVE


```
               [+in]                                    [+in]
orientation  /                        orientation     /
           /   1                                    /   1
fingers  /                            fingers      /
root                                  root
HC                                    HC
Loc/Movement  L    M    L             Loc/Movement  L    M    L
             ipsi     ipsi                         ipsi  str  ipsi
             +ct      +ct                          +ct   +ct  +ct
```
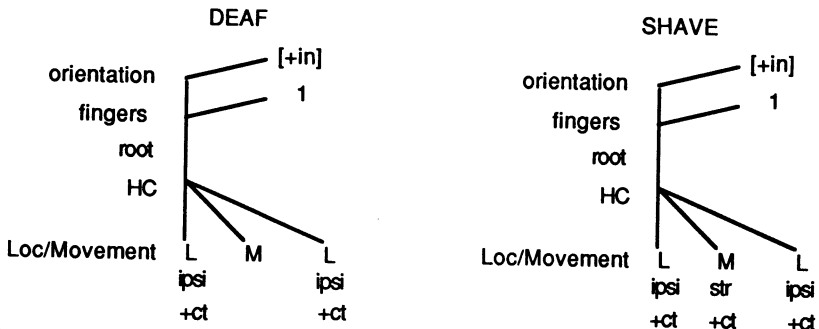
My analysis shows that [-contact] is the unmarked value, and consequently need not be specified in the underlying representations, and, further, predicts exactly which movements will have associated contact.

C. Place Harmony/Metathesis

A more interesting question to ask of the data in (9), and the more difficult one to answer, is why the signs in Group 4, and only Group 4, allow the two points of articulation to be signed in either order. An example is presented below.

14a. FATHER   DEAF
         [POA[forehead]]   [POA[cheek [+top]] < POA [cheek[-top]]]
14 b. MOTHER]   DEAF
         [POA[chin]]   [POA[cheek [-top]] < [POA[cheek[+top]]]

Johnson and Liddell (1989) describe this process as 'metathesis'. They recognize a difference in structure between Group 2 and Group 4, shown below (lower case letters=melody units) :

15a. DEAF   15b. GOOD

(Structure diagrams showing M H M M H over labels a b c d, and H M H over labels e f)
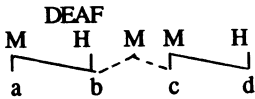
Sandler describes all the groups listed in (9) with the same underlying representation--and LML structure. She does recognize that most monomorphemic signs exhibit what she calls 'Place Harmony', meaning that the same major body area is present in both location segments of such signs (1987b, 1989:135). This is shown below in the sign DEAF. According to Sandler's analysis, Group 3 and Group 4 signs should behave identically, since they have identical underlying representations, yet they do not.

16

DEAF

(Structure diagram with HC connecting to L M L, and [cheek])

[cheek]
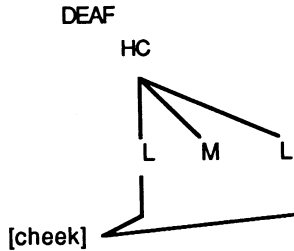
        The differences in behavior between Groups 3 and 4 can be predicted if we utilize the evidence that has been built up in this paper so far, which distinguishes these 4 groups on the basis of underlying structure.
        Group 1 are monosyllabic signs that contain only handshape, place or articulation (POA)and [+contact] specifications. The initial movement to contact is a result of syllabification at W-Level rather than the underlying [direction] feature in SIT, CALL and the other signs in (2).   We know that this is notable difference because of the difference in their behavior in compounds. 'THINK' in THINK-SELF ('decide for yourself') and THINK-DROP ('shocked') has no underlying direction or tracing feature and consequently contributes no moraic unit to the compound, resulting in the monosyllabic compound structure. Recall that the underlying structure for 'CALL/NAME' contains a [direction] feature, resulting in a bisyllabic compound in BLACK-NAME. The result of syllabification in both types of signs is the same, but the underlying representations are different in an important way.   THINK has inserted an unspecified movement inserted in order to satisfy syllabification requirements. The syllabified structures for both THINK and CALL are shown in (17).[11]

17a.   17b.
   THINK   CALL[verb]
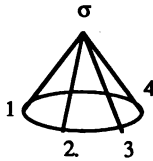
1. [HS[1]]
2. epenthetic syllable Mvt.
3. [+contact]
4. [POA[forehead]]
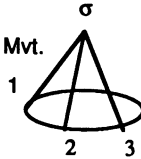
1. [HS[H]]
2. [direction [<--]]
3. [+contact]
4. unspecified POA

Group 2 contains monosyllabic signs that contain 2 contrastive places of articulation underlyingly. The linear order of these two POA values is achieved by virtue of their order on this feature's tier (POA$_1$ < POA$_2$), a formal construct discussed in McCarthy (1989). An unspecified movement will be inserted at P-Level that will make the transition between these two POA's. The syllable structure of GOOD is represented in 18.

18.

GOOD
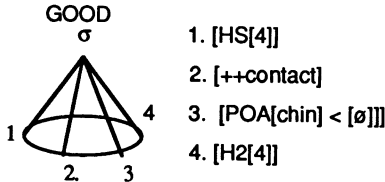
1. [HS[4]]
2. [++contact]
3. [POA[chin] < [ø]]]
4. [H2[4]]

Since Group 2 contains 2 contrastive POA's that are linearly ordered at this level, no metathesis is expected.

Group 3 signs contain an underlying path feature, while Group 4 signs do not. This explains the difference in the distribution of [+contact] as we've seen. If the beginning and ending POA's in Group 3 signs are phonetic consequences of the path, then it is expected that these two points of articulation could not be inverted without changing the underlying path.

Group 4 sign have both 1) a different phonetic and phonological structure than Group 2 signs, and 2) phonetic POA's that are unordered. We must account for the phonetic difference between Group 2 and Group 4 described by Liddell and Johnson (1989). The situation can be accounted for in the following way.

Underlying Level: DEAF [HS[1], [POA[cheek]], [++ contact]
W-Level:
1) Because there is just one underlying POA in Group 4 signs, a syllabic movement is inserted, much in the same way as it is inserted for
   Group 1 signs like THINK and KNOW.
2) Because there is a [++contact] feature value, syllabification proceeds
   to copy the single contrastive POA underlyingly. The only way to
   achieve this double contact at this level is to form a bisyllabic structure of the
   form MPMP.
   The syllabified form is shown in (19).

19.

DEAF
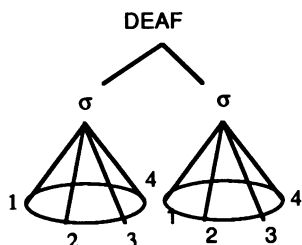
1. epenthetic syllabic Mvt.   1. epenthetic syllabic Mvt.

2. [+contact]   2. [+contact]

3. [[HS[1]]   3. [HS[1]]

4. [POA[cheek]]   4. [POA[cheek]]

P-Level:
1)   An additional movement is inserted between the 2 syllables. This accounts for the slight difference in the shape of the arc, noted by Liddell and Johnson between Group 2 signs and Group 4 signs.
2)   The following P-Level phonotactic constraint applies:

20. **Phonetic Reduplication Constraint**
   Underlying path features may reduplicate in ways that are phonetically identical; underlying POA features may not reduplicate in ways that are phonetically identical; consequently two redundant values are inserted.

(20) accounts for the fact that there are a large number of signs that have identical bisyllabic forms. All of the nominalized signs in (2b) contain two identical syllables.[12] This is because, as Supalla and Newport (1978) originally found and I support, the underlying path is reduplicated in these signs. In the signs that metathesize, however, the lack of an underlying path (in conjunction to the single POA) result in two redundant values.
      In the sign DEAF, the redundant values [cheek[+top]] and [cheek[-top]] are now assigned. What emerges from this account is that it is precisely because the two POA's in DEAF are redundant values, rather than distinctive ones that such signs may metathesize. Further, it is the continuous access to the information at all 3 levels that predicts the nature of these redundant values. As saw earlier in this paper in (1), [±contra] is not specified for all signs. Signs in (1a) will exploit this [±contra] distinction in assigning redundant values. NAVY, CONGRESS, MEMBER, RESTAURANT and FLOWER will have redundant [±contra] features assigned. Signs in (1b) already have a [±contra] specification underlyingly; consequently a 'top/bottom/ [±top] will be assigned in just these cases. DEAF, HEAD, HOME, etc., contain a redundant [±top] value.

III. Conclusion
      From the evidence presented above, we can see that ASL utilizes underspecified underlying representations in its morphophonological and phonological processes. The underlying representation argued for above is significantly reduced from earlier accounts of ASL underlying structure. To sum up, then, we have seen how positing such an underspecified representation:
         1. economizes the compounding process
         2. predicts the distribution of the feature [contact]
         3. predicts the group of signs that will allow metathesis at P-Level.
Moreover, we have seen that such an underspecified representation is important is determining exactly which values can be pressed into service at P-Level as redundant values.

NOTES

[1] Approximate English glosses for ASL signs referred to in the text are typed in upper-case letters.

[2] The place of articulation in the sign SIT can assume a person agreement morpheme. This phenomenon is entirely compatible with this analysis and not the focus of this paper.

[3] [Restrained] is a phonetic phonotactic constraint that doesn't function at W-Level in this analysis.

[4] The verb in the pair is glossed as CALL (As in, "You can call me 'Al'".)

[5] I chose to reproduce Sandler's analysis here because she has explicitly stated that it is the phonology, rather than the phonetic component that she seeks to access. Liddell and Johnson have been working on a 'bottom up' analysis of ASL, and so such a claim is notably absent.

[6] Only Path shape and direction features have sufficient weight in this process to occupy a mora in the phonological syllable(s) of a compound. Syllabic weight has traditionally been equated with sonority (Hyman 1984, Hayes 1989; Perlmutter, ms., for a discussion particular to ASL). Whether or not ASL has a sonority hierarchy is not the subject of this paper; however, this evidence from compounds assists in establishing that there is a notion of syllable weight at work in ASL.

[7] Word-level stress may also play a factor in this process. Compounds containing the sign FACE are exceptions to this: In both FACE-STRONG ('resemble') and FACE-SAME ('look-like' or 'appear-like'), the circular movement in FACE is deleted

[8] My thanks to Carol Padden for pointing out the importance of the compound NAME-SHINE. ('good reputation').

[9] We know that [+contact] is the underlying value rather than [-contact] because :
1. There are a number of signs that are specified [+contact] underlyingly --with [+contact] evident in careful signing--which delete the [+contact] value in fast signing. Some examples are FORGET, THINK, SPACY, UNDERSTAND, SICK, KNOW, RESTAURANT, HEALTH, HEAD, HORSE, and SCOTLAND. This suggests that there may be a widespread phonetic rule that allows the sign to delete contact in such signs.
2. [±distal] signs, such as WANT, DARK, RAINBOW and FLY are redundantly [-contact].

[10] This is one of two types of contact in ASL: 1) contact between the dominant hand (H1) and the non-dominant hand (H2) or body, and 2) contact between selected fingers of H1 and the thumb of H1. For the analysis being presented here, only type 1 is relevant.

[11] At M-Level, the unspecified Path feature in THINK and the unspecified POA feature in CALL/NAME is not represented.

[12] There are also signs that are not nominalized that contain 2 identical syllables, such as COUGH, MILITARY, CANADA. My analysis accounts for these by positing an underlying [direction] feature in these signs, exactly as the nominalized forms.

REFERENCES

Archangeli, Diana. 1984 Underspecification in Yawelmani phonology and morphology. Doctoral dissertation, MIT. [Published 1988, NY: Garland Press]

Goldsmith, John. 1990. Autosegmental and Metrical Phonology. NY: Basil Blackwell, Ltd.

Hayes, Bruce. 1989. Compensatory Lengthening in Moraic Phonology. Linguistic Inquiry 20 (2): 253-306.

Hyman, Larry. 1984. A Theory of Phonological Weight. Providence, RI: Foris.

Kenstowicz, Michael & Charles Kisseberth. 1979. Generative Phonology. NY: Academic Press.

Liddell, Scott & Robert Johnson. 1989. American Sign Language Phonology: The Phonological Base. Sign Language Studies 64: 195-273.

McCarthy, John. 1989. Linear Order in Phonological Representation Linguistic Inquiry 20 (1): 71-99.

Perlmutter, David. ms. A Moraic Theory of American Sign Language Syllable Structure.

Pulleyblank, Douglas. 1988. Vocalic Underspecification in Yoruba. Linguistic Inquiry 19: 233-70.

Sandler, Wendy. 1987a. Sequentiality and Simultaneity in American Sign Language Phonology. Doctoral dissertation. University of Texas, Austin.

Sandler, Wendy. 1987b. Assimilation and Feature Hierarchy in American Sign Language. In Anna Bosch, Barbara Need & Eric Schiller (Eds.) CLS 23 (2): Parasession on Autosegmental and Metrical Phonology, 266-278. Chicago: CLS.

Sandler, Wendy. 1989. Phonological Representation of the Sign. Providence, RI: Foris.

Supalla, Ted & Elissa Newport. 1978. "How Many seats in a chair?" The derivation of nouns and verbs in American Sign Language. In Patricia Siple (Ed.) Understanding Language Through Sign Language Research, 91-133. NY: Academic Press.

Steriade, Donca. 1987. Redundant Values. In Anna Bosch, Barbara Need & Eric Schiller (Eds.) CLS 23 (2): Parasession on Autosegmental and Metrical Phonology, 339-363. Chicago: CLS.

Stokoe, William, Dorothy Casterline & Carl Croneberg. 1965. Dictionary of American Sign Language on Linguistic Principles. Silver Spring, MD: Linstok Press.

GESTURE AND GROUND[1]

Justine Cassell & David McNeill
University of Chicago

I. Introduction

Since the foreground / background distinction was first described (Hopper, 1979; Hopper & Thompson, 1980), it has retained widespread acceptance as a way of analyzing written and spoken texts. Much of the literature since then has taken the contrast more or less for granted and has refined it, changed the terminology, or simply added supporting evidence from additional languages and language families. For English discourse, however, adequate criteria for foregrounding vs. backgrounding have never been described.

We believe that (a) that the dyadic notion of foregrounding / backgrounding misses much of what is going on in narrative texts; (b) part of narrative structure may be carried in the non-verbal channel, as well as the verbal; (c) for English narrative the majority of ground distinctions may be carried by gesture, thus explaining the difficulties encountered in trying to isolate one linguistic system that carries information structure in narrative.

In this paper we will first discuss the distinction between foregrounding and backgrounding as it has been described in the literature. Then we will go on to replace the notion of ground with one of participation frameworks or **event lines** in information, and **narrative level** in speech. And finally we will demonstrate the role that gesture plays in the structuring of narrative.

II. The Concept of Ground

In extended narrative texts, it has been claimed, a distinction is always made between the 'story line' and the 'supporting material'. This distinction has alternately been stated in terms of the material that a narrative is about (sometimes known as the **fabula** in Russian Formalist literature), and in terms of the language in which the narrative is couched (sometimes referred to as the **syuzhet**). That is, Hopper & Thompson (1980) say:

> That part of a discourse which does not immediately and crucially contribute to the speaker's goal, but which merely assists, amplifies, or comments on it, is referred to as **background**. By contrast, the material which supplies the main points of the discourse is known as **foreground**. (1980:280)

and Hopper (1979) says:

> It is evidently a universal of narrative discourse that in any extended text an overt distinction is made between the language of the actual

story line and the language of supportive material which does not itself narrate the main events. (1979:213)

The sort of **material** which is considered foregrounded is events which occur in a temporal order -- that is, punctual non-overlapping events -- and which advance the listener's knowledge of the plot. This resembles Labov's definition of a minimal narrative text as being "a sequence of two clauses which are temporally ordered" (1972: 360). Backgrounded material can be states which do not engage in relations of sequence with events, and which comment on or amplify the main events.

As far as ground distinctions in **language**, the paradigm example of the sort of linguistic distinctions meant is Swahili. Swahili (Hopper, 1979) marks verbs denoting events on the main story line with the prefix *ka-* while events marked as subsidiary receive the prefix *ki-*. Thus, in a story analyzed by Hopper (1979) those events that are punctual, occur in the narrative in the temporal order in which they occurred in the world, and are not simultaneous with or subsumed by other events, are marked with the foreground marker *ka-*.

It is not entirely clear in the work by Hopper, and Hopper & Thompson, whether the concept of ground should be looked for in language or in the story events that exist independently of the discourse which discusses them. In this paper we are going to assume that it is possible to talk about events existing independently of the narrative that talks about them -- although of course these events may be fictive -- and we would like to separate information structure in the story from information structure in the discourse. This issue becomes most necessary when languages without clear morphosyntactic systems for marking information structure are examined.

III. Ground in English

English has repeatedly been the subject of discussions of ground, but never with much success at identifying a stable heuristic for separating foregrounded and backgrounded material.

Hopper & Thompson claimed that transitivity might mark the foreground-background distinction in English and other languages without overt aspectual systems. Their definition of transitivity, however, is not a strictly morphosyntactic one, but also includes semantic contrasts -- that is, it includes a distinction in the kinds of things talked about, and how those things exist in the discourse world. Transitivity involves a number of components: the aspect of the verb, the conscious activity of the agent, and the referentiality and degree of affectedness of the object. For this reason it is impossible to correlate transitivity as a linguistic distinction with a contrast in the kinds of things talked about -- the content criterion for ground. In addition, in practice, Hopper & Thompson's analysis of English texts does not show that transitivity distinguishes ground in this language -- a large percentage of background clauses in the discourse that they analyse have high transitivity values (as mentioned by Delancey, 1987).

IV. Content vs. Linguistic Criteria for Ground

In much of the literature, content criteria for ground have been confused with linguistic criteria (although see Reinhart, 1984). That is, it is unclear whether

what is foregrounded is information that will be referred to by a given clause, or the linguistic nature of the clause itself. In addition, content criteria for ground depend on notions such as 'pivotal events' or 'central and significant information' which contain a covert definition of narrative as concentrating on events in a narrow sense of the term -- excluded are the events of the storytelling, for example.

If, however, we are to look for linguistic devices that express the notion of ground -- which is, after all, the enterprise originally undertaken by Hopper -- then we must have both a non-linguistic distinction and linguistic devices that correlate with that contrast, and the notion of story events (outside of their linguistic instantiation) must be able to handle **all** of the kinds of events narrated in a story.

V. Event Lines

We substitute for the notion of foreground and background, the concept of 'event line' (Cassell & McNeill, in press). The distinction between foregrounding and backgrounding only makes sense in the context of a definition of narrative where single protagonist, single viewpoint events are primary. Events can be seen as foregrounded only if there is one narrator telling one story, and the plot of the **story** is identical to the plot of the **narrating event**. Ground implies a narrating event where the storyteller has only one goal in relating the story. If there are other event chains (looking at a cartoon, recounting the cartoon, representing the organization of the narrative, interacting with the interlocutor, etc), then there cannot be a single or simple distinction between 'foreground' and 'background'. In contrast, we use 'storytelling' or 'narrating' to refer to the entire set of events that make up the conveying of a story by one person to another. Each of these events is grist for the storytelling mill and may be referred to by the storyteller.

We study stories that are told about a cartoon or feature film just seen. For these narrators the following distinctions can be drawn: stories generally refer to what may be called emplotted events: incidents or occurences that follow one another in a real or fictive world (e.g. a character climbs up a drainpipe to reach an upper story).
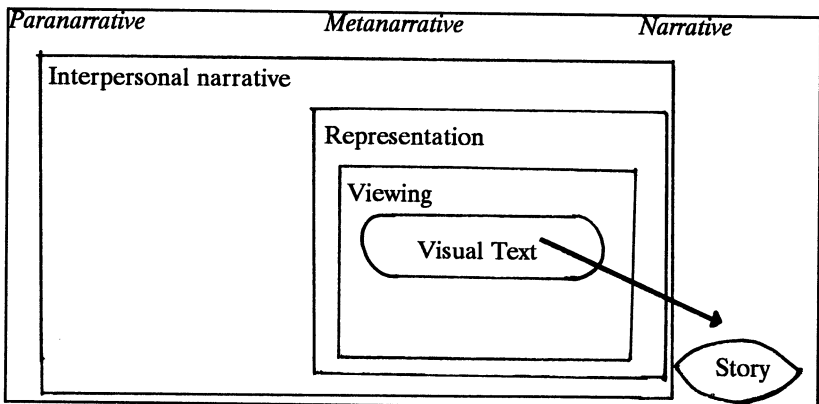
Figure 1: EVENT LINES AND NARRATIVE LEVEL

In addition to emplotted events, though, the stories we are concerned with also commonly contain references to the event of observing the visual text, or cartoon story ('it was a Sylvester and Tweetie cartoon', or 'it was an old movie, a very bad print'), and/or to the event of the storytelling ('I'm going to tell you about a cartoon I just saw'). Reference to these 'metanarrative events' often acts in actual storytelling to create junctures between the parts of a story. The five event lines are listed in Figure 1.

As mentioned above, all of these event lines may be referred to in storytelling. Reference to the visual text, viewing, and speaker's own representation form the **metanarrative** level of the discourse: the part of the narrative that is about narrating. The interpersonal narrative is what we are referring to as the **paranarrative** level of the discourse: the part of the story where the narrator steps out and speaks in his/her own voice to the listener. These notions resemble the three kinds of mimesis described by Ricoeur (1984:52-87).

1. um have you seen any of the uh Bugs Bunny cartoons?	**[PARA]**
{yeah like}	
2. right, ok this one actually wasn't a Bugs Bunny cartoon	**[META]**
3. it was one of the- the series	**[META]**
{oh, ok}	
4. and it had Tweetie Bird and Sylvester	**[META]**
{alright (laugh)}	
5. so so so you know	**[PARA]**
{the cat right?}	
6. right uh huh	
{ok}	
7. and uh the first scene you see is uh	**[META]**
8. this this window with birdwatcher's society underneath it	
9. and there's Sylvester peeking around the window	**[NARR]**

Figure 2: NARRATIVE LEVELS IN SPEECH

They are also listed in Figure 1. The beginning of one speaker's narrative reproduced in Figure 2 gives an example of each narrative level in actual storytelling.

What we call event lines comprise a clear content criterion for ground -- not in terms of the importance of an event but in terms of what aspect of the story is referred to -- what participant frameworks are evoked. Our narrative level, on the other hand, is a way of looking at the linguistic instantiation of ground. It has to do with discourse -- with the clauses that refer to a given event -- and they may be marked in one of a number of ways.

In the next section we are going to talk about how narrative level is marked in gesture.

VI. Gesture and Narrative Level

Narrators make reference to all three levels, and all 5 event lines during a typical story. Adults also demonstrate a typical pattern of gesturing correlated with reference to these different kinds of events (McNeill & Levy, 1982). It is this gesturing which may serve as the indication of ground, or narrative level. We're **not** talking about emblematic gestures (such as those that mean "a-ok", or "V for Victory"). Those emblematic gestures carry meaning in the absence of speech. The four types of gestures that we will describe, on the other hand, occur with speech. They occur most often with the stress peak of a clause, and they carry meaning insofar as they are understood in the context of the speech that they are produced alongside of.

We are claiming that when people tell stories they speak not only of the events of the story *per se* but also of the events of the representation of the story, and of the events of the narrating. And these categories, we claim, can be differentiated by the gestures that accompany them. In Figure 3 is reproduced the same narrative fragment as in Figure 2 but with the gestures noted[2]. We will be making reference to these gestures in the next section.

(A) Gestures at the Narrative Level

The most common kind of gesture produced overall in narratives is the **iconic** gesture. Iconic gestures depict some feature of the action or event being described. Iconics are recognizable by virtue of their form and content. An example is in clause (9) in Figure 3 where the speaker rests her head on her hand and mimes looking left and right while saying "and there's Sylvester peeking around the window. Iconics occur most often with narrative speech. There are several kinds of iconic gestures, corresponding to viewpoint and perspective in narrative, and we will describe these further below.

So, narrative clauses are accompanied by iconics, while meta- and paranarrative clauses are not.

(B) Gestures at the Metanarrative Level

Storytellers tell about the act of narrating by making overt reference to the structure of the story being told, or to the structure of the narrating event. They accompany these references by another kind of gesture, the **metaphoric**. Metaphoric gestures, where the concept being depicted has no physical form, accompany

metanarrative speech, as do deictic gestures. Metaphoric gestures appear with references to the narrative story structure where the cartoon narration may be objectified and commented on in the verbal channel, and presented as an object in the gestural channel. Figure 3 gives an example of a metaphoric gesture: the speaker's hands are rolling over each other while she says, 'it was one of the series.'

```
2. [right], [ok this one] [actually wasn't] a Bugs
Bunny cartoon                                            [META]
     beat           beat            beat

3. [it was one of the- the series]                       [META]
      metaphoric

{oh, ok}

4. and it had [Tweetie Bird and Sylvester]               [META]
                      beat


...

7. and uh [the first scene you see is uh]                [META]
         metaphoric-deictic

8. [this this window] [with birdwatcher's society underneath it]
      iconic                       iconic

9. and [there's Sylvester peeking] around the window     [NARR]
         iconic
```

Figure 3: GESTURE AND NARRATIVE LEVEL

Deictics at this level spatialize, or locate in the physical space in front of the narrator, aspects of the story being narrated. An example in Figure 3 is the hand pointing at empty center space while the speaker says, "the first scene you see is uh".

Another kind of gesture accompanies metanarrative speech and other clauses with a metapragmatic function, and that is the beat gesture. Beats are small baton like movements that do not change in form with the content of the accompanying speech. As well as occurring with metanarrative speech, beats also accompany speech repairs and reported speech. The semiotic value of a beat lies in the fact that it indexes the word or phrase it accompanies as being significant not purely for its semantic content but also for its discourse- pragmatic content. An example in a metanarrative context are, in Figure 3, the beats that accompany each segment in '[right], [ok this one] [actually wasn't a Bugs Bunny cartoon.' Beats can be seen as signalling some kind of move away from the main temporal axis of the story. They reveal, however, that this move can be in the context of a single word -- as in a

repair -- or an entire segment of the discourse -- as for metanarrative speech.

The narrative excerpt in Figure 4 is from the beginning of another cartoon narration by an adult speaker. It demonstrates how gestures function to indicate the kind of background narrative information that is given in the clause they accompany.

The narrator here forms a largish bowl shape with his two hands and raises this bowl from his lap towards the listener. This is a conduit metaphoric gesture: the speaker presents information as if contained in a bounded object that can be passed to the listener (cf. Lakoff & Johnson, 1980; Reddy, 1979). In this example, speech and gesture work together to make clear the metanarrative level at which the narration begins. The speech presents the cartoon as an example of a type, while the gesture represents the narrator, not as an observer, but as the conveyor of this object which is the cartoon. The narrator also performs beats to accompany the introduction of the genre ("it was one of the cat and bird cartoons").

Well ++ [it was one of the ahm Tweetie Pie and ahm] ++ the
                                    beats

cat ++ [cartoons]
              metaphoric: hands present cartoon to listener.

Figure 4: CONDUIT METAPHORIC GESTURE

Conduit gestures like the one just described are not found in narrations in all languages. Whereas we have found excellent examples of conduits in a Georgian language narrative, we have none in Chinese or Swahili narratives. These narratives do contain abundant metaphoric gestures of other kinds, but do not depict abstract ideas as bounded containers. The metaphoric use of space, however, appears in all narratives regardless of the language spoken.

(C) Gestures at the Paranarrative Level

To go on to the third narrative level, when narrators speak as themselves outside of a narrative situation but adopting the role of a participant in a socially defined situation of speaker and hearer, they make only a small number of gestures of a restricted kind. Iconic gestures are virtually absent, as are metaphoric gestures. Deictic gestures are found, to point out the participants of the event: in this case the narrator and listener. This is similar to the findings of Goodwin & Goodwin (to appear), who have observed that gaze plays an important role in structuring the participation of a speaker and hearer in narrating events. Beat gestures are also found, when they mark the **inception** of a paranarrative portion of the discourse, or when they signal repairs or other metalinguistic work. An example of a deictic at this level is pointing towards the listener while saying '[right] uh huh.' Pointing at the listener in paranarrative contexts seems to be a specific clue for this level of narrative organization, opposing it to the narrative and metanarrative levels. The paranarrative point indexes shared knowledge, and the shared activity of narration.

Thus we find a distinctive gestural pattern corresponding to each of the narrative levels: for narrative level, iconics; for metanarrative level, beats, metaphorics, and pointing at empty gesture space; for paranarrative level, pointing

toward the interlocutor.

B. Gestural Layering

Not only do gestures differentiate the three categories of narrative information, but they distinguish within these categories. In Figure 5 are diagrammed the more intricate relationships between gesture and narrative level that will be explained below[3].

```
                                    Narrative level
                                     /    |    \
                               Narr    Metan.      Paran
                                |        |         |
                             Iconic   Metaph   No gesture
                              /         or          or
                            /         Deictic    Deictic
                          /                     (beats signal
                   Focalization                 shifts of level)
                   /         \
               Character    Observer
              /     \      /   \
        Agent     Other /      \
          |         |  | In scene   Outside
          |         |  |  |         |
      Iconic   Iconic  Iconic   Iconic
      C-VPT    C-VPT  O-VPT     O-VPT
              (Gesture acts    (Gesture either
               out specific     surrounds narrator
        character)        or is in front)
```
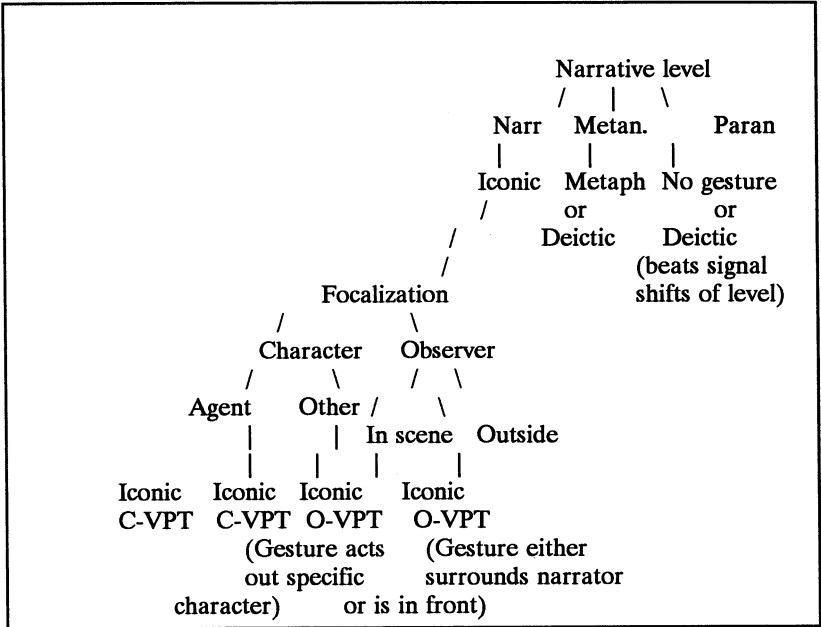
Figure 5: **TYPES OF GESTURES ASSOCIATED WITH NARRATIVE LEVEL**

(A) Layering within the Foreground

Let us go back to the category of story events. Within the narrative level, we find two further distinctions; that of focalization (whose point of view is being taken at the moment), and perspective (where events are viewed from). These distinctions are marked by differences in the kinds of iconic gesture produced.

(a) Focalization

There are two focalizations to consider: a character's or the observer's. A given event may be presented in either one and the alternations are not random wobblings but motivated movements of the speaker closer to or farther from the narrative line. A given event may be portrayed as experienced, or at a distance.

This is manifested directly in the form of the concurrent gestures. The character's focalization narrates events in close-up, as it were, and the C-VPT (character viewpoint) gestures depict this directly. The more peripheral events are narrated at a distance, and the O-VPT (observer viewpoint) gestures present this. Gestural focalization is inferred from the form and space of the iconic gesture. We infer the character as the focalizer when the depiction enacts the character and his movements: the narrator's hand plays the part of a character's hand, her body the part of this character's body, etc. The gesture space includes the speaker's own body as an integral part. An example from Figure 3 is the gesture previously discussed: "and there's Sylvester peeking around the window"

Conversely, an O-VPT gesture concentrates the entire event in the speaker's hand. The hand equals the entire character, its movement is the motion of this character, not just one part of him, and the gesture space is localized in front of the speaker, excluding her own body. The speaker is still part of this gesture, but now as an observer of the event. An example is given in Figure 6.

"he tries going [up the inside of the drainpipe and]"
> O-VPT iconic: right hand forms a fist with forefinger extended and moves straight up. Narrator's hand depicts the character's body.

Figure 6: OBSERVER FOCALIZATION

We may use gesture as a clue to the function of accompanying speech. The character voice tends to appear with single-clause sentences that use, where possible, active, transitive verbs and minimal subjects: An illustration is 'and ∅/ drops it down the drainpipe' which was accompanied by a C-VPT gesture. The O-VPT conversely tends to appear with multiple-clause sentences. Corresponding to the distance built into the O-VPT gesture, distance is introduced in these sentences since one clause describes the action and the other clause may express a narrative observer. Another example with an O-VPT gesture is, 'and we see him go up the drainpipe,' where one clause indexes the observer's focalization. Thus, even though in English the narrative foreground is said to be not specifically indexed linguistically, one kind of layering occurs within the narrative foreground in gestural form, and this leads us to note an unsuspected parallel in syntax.

(b) Perspective

The observer may stand either outside the event as an onlooker on the event, or may stand inside the event but not as a participant in the action. An example is using a hand to show a figure swinging past the narrator herself, from back to front, while saying, 'and you see him swinging down [across a rope].' Unlike a C-VPT gesture, the gesture depicts the character as a whole entity, not just the hands of the character. However, as in the C-VPT, the gesture space extends backward to include the narrator. In an outside perspective, in contrast, the space is in front of the narrator. We have not found parallels between gestural perspective and the syntax of accompanying speech. This gesture is shown in Figure 7.

```
1. Internal Perspective:

"and you see him swinging down [across a rope]"
        iconic: fist swings from behind head towards listener

2. External Perspective:

"and he puts a [board on a box]"
        both hands describe box and then board in center space
```

Figure 7: PERSPECTIVE WITHIN NARRATIVE

Within the foreground, then, we find a rich structure of subdivisions. The speaker's sense of the perceived centrality vs. peripherality of the events in a narrative takes on concrete form in gesture focalization and perspective. Narrative distance may be represented by the actual physical distance of a narrator from the narrated. These images do not only come out of, but also can have real ongoing consequences for the speaker's thoughts, and therefore for narrative and memory. Thus, for both the speaker, and for the listener, gestures help to build a representation of the narration, at all of its levels, and play an important part in the 'telemention' of the story (Harris & Taylor, 1989).

(B) Layering within 'Extranarrative'

As mentioned above, within the background we find a distinction between reference to the structure of the story (metanarrative), and reference to the structure of the narrative event itself (paranarrative). We can see this distinction as another example of perspective in gesture.

```
A. METANARRATIVE

"and it had [Tweetie Bird and Sylvester]"
        metaphoric: indicates left then right to spatialize characters


"and uh [the first scene you see is uh]"
        deictic: indicates center as locale for new scene

B. PARANARRATIVE

"[right] it's one of the series"
        deictic: point at listener

"[so so so you know]"
        deictic: point at listener
```

Figure 8: PERSPECTIVE WITHIN NON-NARRATIVE

Although the same space may be pointed out in gesture, depending on the narrative level of the accompanying speech, the gesture will have a different value.

Thus in example (A), Figure 8, the first deictic spatializes an opposition among characters, the second deictic spatializes an opposition among scenes. At the paranarrative level, on the other hand, as under (B) in Figure 8, the deictic spatializes an opposition between speaker and hearer: more exactly it indexes both of them as participating in the same speech event.

V. Conclusions

In this paper we have argued for the following major conclusions: (a) the theory of ground contains a covert naive definition of story that conflates the different kinds of events, and different kinds of goals that a speaker may talk about. (b) ground is not capable of describing all of the narrative phenomena that have to do with the relationships between different kinds of events in narrative. (c) narrative structure, which has seemed problematic for English speech, is more clearly marked by nonverbal means. Once we can discover from gestures where the background and foreground are distinguished, we can sometimes observe hitherto unsuspected linguistic contrasts, as in the case of simplex/complex sentence structure being used for different focalizations. Observation of the gestures that accompany narrative discourse illustrates the potential for non-spoken resources of narrative structure that in English may be richer than those available to speakers in the verbal channel.

Footnotes

2. Square brackets mark the extent of a gesture. Plus signs (++) indicate the duration of a pause.

3. C-VPT stands for 'character viewpoint' and O-VPT stands for 'observer viewpoint'.

Bibliography

Cassell, Justine & David McNeill. (in press). Gesture and the Poesis of Prose. Poetics Today.

Delancey, Scott. 1987. Transitivity in Grammar and Cognition. Coherence and Grounding in Discourse. Amsterdam: John Benjamins. pp.53-68.

Goodwin, Charles and Marjorie Harness Goodwin. (in press). Context, Activity and Participation. The Contextualization of Language, ed. by Peter Auer & Aldo di Luzo. Amsterdam: Benjamins.

Harris, R. & T.J. Taylor. 1989. Landmarks in Linguistic Thought: the Western Tradition from Socrates to Saussure. London: Routledge.

Hopper, Paul. 1979. Aspect and Foregrounding in Discourse. Syntax and Semantics (Vol. 12): Discourse and Syntax. New York: Academic Press. pp. 213-241.

Hopper, Paul & Sandra Thompson. 1980. Transitivity in Grammar and Discourse. Language Vol 56:2. pp. 251-299.

Labov, William. 1972. The Transformation of Experience in Narrative Syntax. Language in the Inner City. Philadelphia: University of Pennsylvania Press.

Lakoff, George & Mark Johnson. 1980. Metaphors We Live By. Chicago: University of Chicago Press.

McNeill, D. & Levy, E. (1982). Conceptual Representations in Language Activity and Gesture. Speech, Place, and Action, ed. by R. Jarvella & W. Klein. Chichester: John Wiley & Sons.

Reddy, Mark. 1979. The Conduit Metaphor -- A Case of Frame Conflict in our Language about Language. Metaphor and Thought, ed. by A. Ortony. Cambridge: Cambridge University Press. pp.284-324.

Reinhart, Tanya. 1980. Principles of Gestalt Perception in the Temporal Organization of Narrative Texts. Linguistics 22. pp.779-809.

Ricoeur, Paul. 1984. Time and Narrative (vol. 1). Translated by Kathleen McLaughlin & David Pellauer. Chicago: University of Chicago Press

The Relevance of Syllable Structure in Place Assimilation

Young-mee Yu Cho
Stanford University

0. The relationship between syllable structure and assimilation has not been explored in full in the generative phonology. In the literature there has been a common assumption that the coda consonant assimilates to the onset consonant because the weak coda position is dominated by the strong onset position (Vennemann 1972, Hooper 1976, Murray 1982). In this paper, I argue against that claim and propose that the coda-target condition (i.e. the target being in the coda) should be identified as one essential condition in place assimilation. I am going to illustrate that coda consonants assimilate (but not necessarily to onset consonants) when the other conditions are met.

First, I will present two cases discussed in the literature, and then argue that a stronger case can be made for Sanskrit, in which the target's being in the coda, though not a sufficient one, is a necessary condition in predicting all and only the types of assimilation in the language.

1. Harris (1969) analyzes Spanish homorganic nasal assimilation in terms of two rules to account for the data in (1): first, nasals assimilate to all consonants other than glides within and across word boundaries, and second, nasals assimilate to glides only across word-boundaries.[1]

(1) Spanish Nasal Assimilation (Harris 1969, Hooper 1976, Penny 1986)

Word-Internal	Across word boundaries
campo 'field'	son [m] perros 'they are dogs'
canto 'I sing'	son [n] tigres 'they are tigers'
baŋko 'bank'	son [ŋ] carros 'they are cars'
enfriar [mf] 'to cool'	han [m] fregado 'they have scoured'
ensillar [ns] 'to saddle'	han [n] sido 'the have been'
nieto [njeto] 'grandson'	un hielo [uñ jelo] 'a piece of ice'
nuevo [nweBo] 'new'	un huevo [uŋ weBo] 'an egg'

Since glides trigger assimilation only across word-boundaries, Harris (1969) was forced to posit two rules. In his two-rule approach, glides are excluded from triggering assimilation within words by stipulating that only consonants with the feature [+consonantal] act as triggers within words, whereas across words any consonant can trigger the rule. Hooper (1976) observes that Nasal Assimilation can be viewed as a single rule if one incorporates the notion of

syllable structure. In particular, she argues that the rule applies when the nasal and the consonant are in different syllables. In *nieto* and *nuevo* there is no assimilation because the target and the trigger are tautosyllabic. Since the consonant in the stronger position influences the consonant in the weaker position, Nasal assimilation in Spanish occurs before the glides (y and w), but only when the nasal is syllable-final and the glide is syllable-initial (Penny 1986). Spanish illustrates that assimilation rules are somehow related to syllable structure because specifying the directionality of the rule application as from right to left is not sufficient to distinguish between the different behavior of the two kinds of glides. However, it does not, in itself, argue for the coda-target condition since it is possible to formulate the rule as applying either when the target and the trigger are heterosyllabic, or one can even say that the trigger has to be in the syllable-initial position.

Place assimilation in Central Catalan (Mascaró 1987) shows that the target of place assimilation should be characterized as occupying the coda, though the trigger does not have to be in the onset. As illustrated in the examples in (2), there is regressive assimilation, which appears to spread the place features of the onset of the following syllable to the preceeding coda consonant when the coda consonant is dental. The last example shows that this apparent leftward spread is in fact a case of coda target: in *añʃ*, /s/ assimilates to /ñ/ even though the trigger precedes the target and the trigger is not in the onset.

(2) Catalan Place Assimilation (Mascaró 1987)

> set beus → sebbeus 'seven voices'
> set kosins → sekkosins 'seven cousins'
> son beus → sombeus 'they are voices.'
> son kosins → soŋ kosins 'they are cousins.'
>
> añ+ s → añʃ 'years'

Another case confirming the relevance of the coda target condition is Korean, in which in some rare coda clusters, a dental assimilates to a neighboring coda consonant progressively (Cho 1988). The fact that we normally find regressive assimilation (i.e. the right-to-left directionality) is, then, due to the fortuitous surface fact that there is usually one coda consonant and one onset consonant.

2. I make the following theoretical assumptions. First, in Autosegmental Phonology rules of assimilation involve not a change or a copy but a reassociation of the features of one segment, so that they come to include the other segment in their scope. This operation of reassociation is called spreading as shown in (3), and it is the sole mechanism of assimilation rules (Goldsmith 1979, Steriade 1982, Hayes 1986).

(3) Assimilation as Spreading

(any node in Feature Geometry)

F
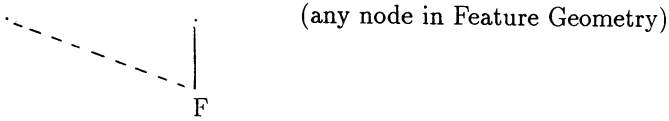
Another often-noted generalization regarding assimilation is the asymmetry between possible propagating values; i.e. assimilation tends to spread marked feature specification to segments. This asymmetry finds its natural explanation in Underspecification Theory which requires marked values to be specified and unmarked values unspecified (Kiparsky 1982, Archangeli and Pulleyblank forthcoming). I assume Radical Underspecification where every predictable feature specification is absent in underlying representation so that there is a stage in which only marked feature specifications are present and play a role in assimilation.

Another theoretical assumption that is crucial in predicting only the kinds of assimilation that actually occur is the principle of Structure Preservation (Kiparsky 1985, Borowsky 1986), which is expressed in terms of constraints that apply in underlying representations and to each stage in the derivation up to the level at which they are turned off.

Also I assume that segments are not unordered feature bundles but that they are hierarchically organized, following Clements (1985) Sagey (1986) and McCarthy (1988). I will adopt the representation of the Place Node for Sanskrit, as shown in (4).

(4) Place

.Labial . Coronal . Dorsal

[round] [ant] [dis] [lat] [high]
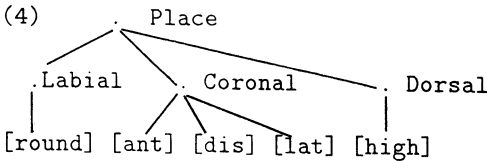
Lastly, following Steriade (1987), I assume that assimilations result from setting up the values of a very limited set of universal parameters, such as the site of spreading, the locality condition, the specification on the target and (or) trigger, and directionality.

3. Sanskrit Coronal Assimilation comprises the retroflexion, palatalization and lateralization of dentals. Although the three processes appear unrelated on the surface. I will argue that, once the coda target condition is adopted in place assimilation, they can be unified as one rule which spreads the Coronal Node, rather than individual features.

(5) shows the coronal consonants of Sanskrit.

(5) Sanskrit Coronal Consonants

	dental	retrofl	palatal
stops	t	ṭ	c
	d	ḍ	j
	n	ṇ	
liquids	l	r	
fricatives	s	ṣ	ś

Focussing on the Coronal Node, I will assume that the features [anterior], [distributed] and [lateral] are its dependents (Levin (1987)) in order to distinguish the four types of coronal consonants (dental, retroflex, palatal and lateral). (6b) lists all the features assumed to be present underlyingly. For instance, /t/ will not be marked for any feature specification, and /n/ will be marked only for the feature [+nas] though it has no specification of place features.

(6) a. Feature Specifications
  dentals: +ant, –dis
  retroflexes: –ant, –dis
  palatals: –ant, +dis
  laterals: +lat, +ant, –dis
 b. Underlying Feature Specifications
  dentals :   `
  retroflexes: –ant, –dis
  palatals: –ant
  laterals: +lat

Under the assumption that only specified features spread, it is not surprising that a dental becomes a palatal and a lateral in the right environment. I will discuss palatalization and lateralization first, since they appear to be considerably more regular than retroflexion. As illustrated in (7), all dental consonants *t, s, n* assimilate to whatever palatal consonant follows them,[2] and *l* and *n* assimilate to a following lateral.[3]

(7) Palatal/ Lateral Assimilation

mahān + kaviḥ →	mahānkaviḥ	'great poet'
mahān + bhāgaḥ →	mahānbhāgaḥ	'illustrious'

tān + janān →	tāñjanān	'those people'
trin + lokān →	trĭllokān	'three worlds'

ut + carati →	uccarati	'rise'
etat + chattram →	etacchattram	'this umbrella'
vidyut + jāyate →	vidyujjāyate	'the dawn is born'
tat + labhate →	tallabhate	'it takes'

tatas + ca →	tataś ca	'and then'

At this point, the Palatal/Lateral Assimilation can be tentatively formulated as spreading the relevant feature, [−ant] and [+lat] respectively, as shown in (8).[4]

(8) Palatalization/Lateralization

```
  .        . (Place)           .        . (Place)
  |        |                   |        |
  |        |                   |        |
  |        | (Coronal)         |        | (Coronal)
  .  ~  ~  |                   .  ~  ~  |
        ~  |                         ~  |
       [-ant]                      [+lat]

 (Palatalization)              (Lateralization)
```
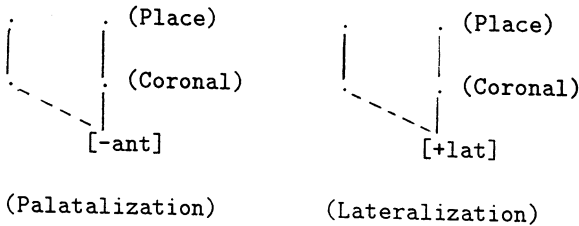
Now let us discuss Retroflex Assimilation. On the surface, Retroflex Assimilation seems to behave quite differently from lateralization and palatalization. Some relevant examples are listed in (9a), and (9b) is a table which shows what happens to the potential targets, which is represented by the first column, when they are followed by the potential triggers in the first row. 'Yes' represents assimilation while 'no' represents non-assimilation. Some of the potential targets undergo different rules which bleed the assimilation rule.

(9) a. Retroflex Assimilation

tān + ḍimbhān →	tāṇḍimbhān	'those infants'
tat + ḍhaukate →	taḍḍhaukate	'it approaches'
tat + ṭīkā →	taṭṭīkā	'the commentary about that'
pādas + ṭalati →	pādaṣṭalati	'the foot is disturbed'
pātas + ṣaṭ →	pādaṣṣaṭ	'the foot six'

b. target\trigger	ṭ	ḍ	n	ṣ	r
t,d	yes	yes	yes	no	no
n	mṣṭ	yes	yes	mṣ, ntṣ	mr
s	yes	rḍ	rṇ	yes	no

Stops (both oral and nasal) always trigger assimilation of the preceding segment, be it a fricative, a stop, or a nasal. On the other hand, the voiceless fricative /ṣ/ triggers assimilation only when the preceding segment is also a fricative. When the preceding consonant is a stop, there is no assimilation.[5] Another peculiarity of triggers concerns /r/, which never triggers assimilation.[6]

In spite of the apparent irregularities I will propose that retroflexion behaves exactly like the other types of coronal assimilation if we assume that the target has to be in the coda. If we do not incorporate the coda target condition, such rules as in (10) could be suggested. (10a) is the case when stops are triggers, whereas (10b) represents the fact that the fricative /ṣ/ is not as permissive as the other triggers. The fricative retroflex behaves as the trigger only when the target is the fricative dental /s/.

(10) Retroflex Assimilation

a)
```
  .         .
  |         |     [-cont]
  |         |     (Place)
  .         .
  |         |     (Coronal)
  .\        .
    \     |
   [-ant, -dis]
```

b)
```
            .   .────────────
 [+cont]|   |   |  [+cont] [-voice]
        |   |   |     (Place)
        .   .   .
        |       |     (Coronal)
        .\      .
          \   |
         [-ant, -dis]
```
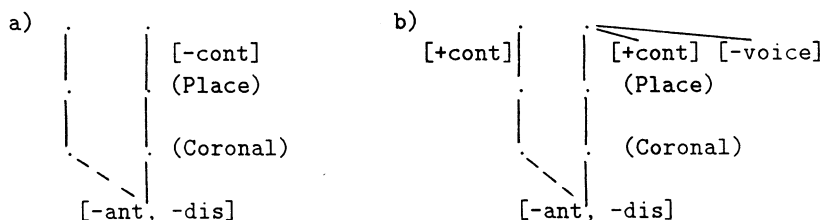
(when the trigger is /t/,/n/)   (when the trigger is /ṣ/)

We immediately note several problems with the above rules. First, we predict with no justification that these are two independent rules and that they can be manipulated separately. Second, we need to stipulate a condition on the target and the trigger that they have to share the feature [+cont] for the fricative trigger. I have argued in Cho (1989) against the so-called 'the principle of similarity,' which states that assimilation rules tend to apply when certain features are shared between the target and the trigger. Given that such a condition is not only undesirable on formal grounds but also empirically not supported, (10b) involves a stipulative complication.

The third and the most serious problem is; whereas all instances of spreading have been dealt with at the point in derivation where none of the default rules have been specified, for these rules we need a stage in which such default values as [-voice] and [-cont] have been filled in. In order to guarantee that

only dentals are the targets, we have to rely on a stage in which only the features that characterize the retroflex, lateral and palatal consonants are present and referred to as a set of spreading features.

I argue that all the surface peculiarities concerning the fricative trigger can be given a natural explanation by assuming independently motivated syllable structure. A syllable-based account (i.e. the coda-target condition) posits only one coronal assimilation which accounts for 4 different rules of (8) and (10). Nothing special has to be said about the nature of the target or the trigger with regard to retroflexion.

There are three sets of consequences for potential targets. One is the case in which a universal sonority scale (as shown in (11a)) predicts whether a sequence of two consonants should be tautosyllabic or heterosyllabic. Following Steriade (1982), I assume that the rule which forms onsets precedes the coda rule. For instance, an oral stop is always equal to or less sonorous than any consonant and has to syllabify as the onset, thereby triggering assimilation in all cases.

(11) Syllabification

　a. Sonority
　stops　　fricatives　　nasals　　liquids　　glides
　————————————————————→ (more sonorous)

　b. Negative Condition on onsets
　*[tn　　*[tl　　*[nl
　σ　　　σ　　　σ

　c. Syllabification
　　s] [ṣ　　t] [t̪　　s] [t̪　　t] [l　　t] [n
　　[tṣ　　[tr　　[sr

　d. Three Arguments for Syllabification (Steriade 1982, Varma 1961)
　1) Reduplication: ta-sthau
　2) the Aorist /s/ deletion: /a-chid-s-ta/ → [achitta]
　3) Doubling: /hasta/ → [hastta]

The second is the case in which syllabification is governed not by sonority but by a language-particular condition: certain homorganic clusters are not allowed as onset clusters as reresented in (11b). The third is the case in which, due to prior application of neutralization, certain clusters (/nt̪/, /nr/ and /sr/) never arise, thereby bleeding the assimilation rule in question.

Let us now see in detail how the proposed coda-target condition renders Retroflex assimilation completely regular. Under the new account, /ṣ/ triggers assimilation to a fricative not because both the target and the trigger are continuants, but because, as shown in (11c) a sequence of two fricatives always syllabify as heterosyllabic, syllabifying the first fricative to the coda position. A sequence of oral stop and fricative, however, is always tautosyllabic since stops

are less sonorous than fricatives. Since *t* does not occupy the coda position, there is no assimilation in the *ts̩* cluster. In this account, one need not specify either the target or the trigger, other than that the target has to be in the coda; i.e. any retroflex consonant triggers assimilation only when the target is in the syllable coda position. (11c) represents syllabification facts of some relevant clusters. If we look at such cases as s̩s̩, t̩t̩, s̩t̩, ṉḍ, etc. in (10b), in which assimilation takes place, the first consonant is always equal to or more sonorous than the second consonant and has to syllabify as the coda of the preceding syllable.

The same explanation can be given for the special behavior of /r/. The potential target (the obstruent that precedes /r/) is never in the coda position since any obstruent is less sonorous than /r/, and syllabifies in the same syllable as /r/.[7]

Let us now look at the /s-t/ cluster in detail. According to Steriade there are two arguments for the /s/ + stop cluster to be heterosyllabic, even though there are word-initial /s/ + stop clusters. One is the fact that *s* does not participate in reduplication which involves copying of the root melodic core, as exemplified in *ta-sthau*. The other is the deletion of the Aorist suffix *s* which deletes when flanked by less sonorous consonants. As illustrated in an example like /a-chid-s-ta/ → [achitta], *s* cannot be syllabified with either the preceding or the following *t*. The third piece of evidence is a doubling rule which I analyze as the gemination of the first onset consonant. In the /s/+ stop cluster, it is not the *s* that doubles but the following stop as in /hasta/ → [hastta].

If syllabification takes place only in terms of the relative sonority strength of the two consonants, one wonders why there is assimilation in *tn* and *tl*, in which the nasal or the liquid is more sonorous than the stop, and both of the segments should be tautosyllabic. In these particular sequences, there is yet another independently needed onset constraint at work. Even though the sonority scale licenses such sequences as tautosyllabic, Sanskrit, like many other languages, does not allow two homorganic sequences as onset clusters, thereby prohibiting such clusters as *tn, tl, pm*, etc. Due to this negative condition, any homorganic clusters should be heterosyllabic. Given that *t* in *tn̩* and in *tl* occupies the coda position, we can explain why the nasal and the lateral features spread to it.

The third, and the last case that requires explanation is the case in which one expects assimilation based on syllabification but there is no assimilation. In the sequence *ns̩*, the sonority scale would force *n* to be syllabified as the coda of the preceding syllable. However, in internal sandhi there is a rule that derives a nasal glide known as anusvāra from a nasal stop whenever it precedes a continuant and this rule bleeds Retroflex Assimilation.

In external sandhi, there is always an intrusive stop /t/ between the nasal-

fricative sequences like *nṣ, ns* → *ntṣ, nts*. Whitney (1889) observes that the insertion of the /t/ in such cases is a necessary operation. Schein and Steriade (1986) treat these sequences of nasal and stop as underlying clusters in the final position rather than as having an intrusive stop. Whichever analysis one might choose for the nasal-stop cluster in the final position, we can account for the reason why there is no assimilation.

After the sequence *nṣ* has developed into *ntṣ*, we expect the syllabic division to be *nStṣ* rather than *ntSṣ* due to the fact that the onset rule precedes the coda rule. First, *t* is not in the coda position and cannot undergo assimilation. As for the coda *n*, even though it occupies the coda position, the locality condition which requires the target and the trigger to be skeletally adjacent prohibits the retroflex feature from spreading from *ṣ* to *n* crossing the skeletal position occupied by *t*.

It should be emphasized at this point that Patalization and Lateralization behave in a completely regular fashion due to the fortuitous fact that all the potential targets happen to be in the coda position. On the other hand, Retroflexion appears to be irregular since some of the potential targets are in the onset. It is quite significant that syllable structure which is needed on indepedent grounds can be consistently applied as a condition in place assimilation. Now the parameters involved in coronal assimilation are shown in (12).

(12) Parameters for Coronal Spread
    Site of Spreading: the Coronal Node
    Target Specification: Coda target
    Rule Order: Spread precedes all redundancy rules.
    Locality Condition: skeletal adjacency
    Directionality: none
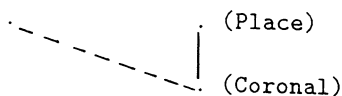
4. I conclude that it is more than a coincidence that a potential target emerges as an actual target just in the case it occupies the coda position. Incorporating the coda-target condition can not only unify 4 otherwise unrelated rules (one for palatalization, one for lateralization, and two for retroflexion) as a single rule that spreads the CORONAL node, but it also helps to clarify the role of syllable structure in assimilation.

The coda target condition clearly shows that assimilation processes, though commonly attributed to phonetic explanations, are phonologically governed. Ohala (1988), among others, offered phonetic explanations for the common types of place assimilation. He claimed that place assimilation is not due to

the speaker's innovation but due to acoustic-auditory reasons; namely, the burst at release provides a more robust cue to place of articulation than the formant transitions that occur between two articulations. These phonetic explanations may provide some answers as to why real time changes work the way they do but these are not sufficient to explain how a particular synchronic grammar works. As a result, they do not have a direct bearing on the types of assimilation that actually occur and the conditions under which assimilation occurs. The phonetic explanations are valid only to the extent that they predict why the realization of the parameters are more likely to be one way rather than the other. That is, it can explain why Coda-Target is a more plausible value for the target-specification parameter than Onset-Target.

In the phonology, it is not directly relevant whether there is [release], or phonetic cues in the language or not. If the non-release (thus not providing relevant place cues) were a direct cause of place assimilation, we would not expect cases like Korean and Catalan where there are more than two consonants in the coda and the second consonant, though released, assimilates to the first, which is unreleased. Nor does it handle the complex case of Sanskrit retroflexion. I believe the coda condition is yet another instance showing that phonologization cannot be deduced from acoustic/ auditory organizations (Hyman 1976) and also that phonology is not natural (Anderson 1981).

Notes

* I would like to express my thanks to G. Guy, M. Inman, P. Kiparsky, A. Lahiri, W. Leben for comments and advice.

[1] Note that in Spanish [n] is the only nasal permitted word-finally due to a neutralization rule (Harris 1984).

[2] When *t* is followed by a palatal fricative *ś*, there is a mutual assimilation, so that the sequence of *t-ś* results in *c-ch* as in *vedavic chuurah* and *tac chrutvaa*.

[3] In external sandhi, *n* assimilates only to the coronal place of articulation whereas in internal sandhi it assimilates to all places of articulation. Schein and Steriade (1986) attribute the behavior of word-final *n* to the fact that the word-final *n* contains underlying clusters as in *-ns* (Acc. pl) and *-nt* (3pl).

[4] I assume that the reason /s/ does not assimilate to the lateral is due to Structure preservation. *[+cont, +lat] is not possible whereas [+son, +lat] is.

[5] In internal sandhi /s/ changes according to the environment. According to Whitney, the only environment in which /s/ remains unchanged is before /t/ and /th/.

[6] We should note that /r/ is a trigger for the ruki rule which retroflexes *s*. However, the ruki triggers, that is, *r, u, k, i* do not seem to be characterized by the features dominated by the Coronal Node.

[7] For some reason, the ṇt cluster is realized as mṣṭ rather than ṇṭ. Odden

(1978) proposes an *s*-epenthesis rule. Note that retroflex assimilation takes place even here because *s* is in the coda and realized as a retroflex fricative even though the nasal part is realized as a nasal glide known as anusvāra.

References

Allen, W. S. 1962. *Sandhi.* Mouton: the Hague.

Anderson, S. 1981. "Why phonology isn't natural." *Linguistic Inquiry.* 12: 493-539.

Borowsky, T. 1986. *Topics in the Lexical Phonology of English* Doctoral Dissertation. University of Massachusetts.

Archangeli, D. and D. Pulleyblank. forthcoming. *The Content and the Structure of Phonological Representations.* Cambridge, MA: MIT press.

Cho, Y. Y. 1988. "Korean Assimilation." *Proceedings of the Seventh West Coast Conference on Formal Linguistics.* vol. 7. 41-52.

Cho, Y. Y. 1989. "On the principle of similarity in consonantal assimilation." *Proceedings of CLS.* 25.

Cho, Y. Y. 1990. *The Parameters of Consonantal Assimilation.* Doctoral dissertation. Stanford University. Stanford, California.

Cho, Y. Y. forthcoming a. "On the Universality of the Coronal Node." C. Paradis and J.F. Prunet eds. *The Special Status of Coronals .* Academic Press.

Clements, G. N. 1985. "The geometry of phonological features." *Phonology Yearbook.* 2. 223-250.

Clements, G. N. and S. J. Keyser 1983. *CV Phonology* Cambridge, MA. MIT Press.

Goldsmith, J. 1979. "The aims of the autosegmental phonology." In Dinnsen, D. A. ed. *Current Approaches to Phonological Theory.* 202-222.

Harris, J. 1969, *Spanish Phonology.* Cambridge, MA. MIT Press.

Harris, J. 1984. "Autosegmental Phonology, Lexical Phonology, and Spanish Nasals." M. Aronoff and R.T. Oehrle eds. *Language Sound Structure.* Cambridge, MA. MIT Press. 67-82

Hayes, B. 1986. "Assimilation as Spreading in Toba Batak." *Lingusitic Inquiry.* 17. 467-99.

Hooper, J. 1976. *An Introduction to Natural Generative Phonology.* New York. Academy Press.

Hyman, L. 1976. "Phonologization." In A. Julliand (ed.) *Linguistic Studies Offered to Joseph Greenberg.* 407-18. Saratoga, CA: Alma Libri.

Kiparsky, P. 1982. "Lexical Morphology and Phonology." In Yang, I. (ed.) *Linguistics in the Morning Calm.* Seoul. Hanshin.

Kiparsky, P. 1985. "Some Consequences of Lexical Phonology." *Phonology Yearbook.* 2: 85-138.

Levin, J. 1987. "A Place for Lateral in the Feature Geometry." Paper presented at 1987 LSA Annual Meeting.

McCarthy, J. 1988. "Feature geometry and dependency." *Phonetica.* 43: 84-108.

Mascaró, J. 1987. "A reduction and spreading theory of voicing and other sound effects." ms. Universitat Autonoma de Barcelona.

Murray, R. W. 1982 "Consonant Cluster developments in Pali." *Folia Linguistica Historica.* 3. 163-184.

Odden, D. 1978. "Further evidence for the feature [grave]." Linguistic Inquiry 9. 141-44.

Ohala, J. 1988 "The phonetics and phonology of aspects of assimilation." Paper presented at the 1st Conference on Laboratory Phonology, Ohio State University. To appear in M Beckman and J. Kingston eds.

Penny, R. 1986. "Sandhi phenomena in Castilian and related dialects." In H. Andersen ed. *Sandhi Phenomena in the Languages of Europe.* Mouton de Gruyter. Amsterdam. 505-518.

Pulleyblank, D. 1986. *Tone in Lexical Phonology,* NLLT Collection. Reidel.

Sagey, E. 1986. *The Representation of Features and Relations in Non-linear Phonology.* Doctoral Dissertation. MIT. Cambridge, Massachusetts.

Schein, B. and D. Steriade. 1986. "On Geminates," *Linguistic Inquiry.* 17. 691-744.

Steriade, D. 1982. *Greek Prosodies and the Nature of Syllabification,* Doctoral Dissertation. MIT. Cambridge, Massachusetts.

Steriade, D. 1987. "Locality Conditions and Feature Geometry." *NELS.*

Varma, S. 1961. *Critical Studies in the Phonetic Observation of Indian Grammarians.* Royal Asiatic Society. London.

Vennemann, T. 1972. "On the theory of Syllabic Phonology" *Linguistische Berichte.* 18. 1-18.

Whitney, W. D. 1889. *Sanskrit Grammar.* Harvard University Press, Cambridge, Massachusetts.

Experiential vs. Agentive Constructions in Korean Narrative*

Soon Ae Chun and David A. Zubin
State University of New York at Buffalo

1. Introduction. There are parallel syntactic constructions in Korean, depicted in (1a) and (1b), into which a variety of perception and psych verbs can be inserted.1

(1) a. *Suni-ka/nun   Toli-lul   coahanta.*
       Suni-NOM/TOP Toli-ACC like   'Suni likes Toli.'

  b. *Suni-eykey/ka/nun   Toli-ka   coh-ta.*
     Suni-DAT/NOM/TOP Toli-NOM likable   'Toli is likable to Suni.'2

Both versions contain an experiencer and an object of experience, but the case marking differs. In (1a) the experiencer is nominative and the object of experience, or stimulus, is accusative. In (1b) the experiencer is dative or nominative and the stimulus is nominative. We will call sentences like (1a) the agentive construction and those like (1b) the experiential construction. While various syntactic characterizations have been proposed for the experiential construction,3 its inherent semantics and how they differ from those of the agentive construction have been neglected. Either it is derived from the agentive construction, suggesting that the two are synonymous, or a more or less fortuitous semantic feature is attached to the experiential construction.4

    This paper will present an analysis of agentive and experiential constructions based on differences that can be explained by the notion of subjectivity, a property of cognition that underlies all language use. Section 2 is an introduction to subjectivity. Section 3 presents the morpho-syntactic relations between these two constructions, demonstrating a productive structural opposition, akin to voice, in the grammar of Korean. Sections 4 and 5 present the distributional effects that are brought about by the expression of different degrees of subjectivity in interactional discourse and in narrative respectively. Finally in Section 6, the cognitive folk models that underly the two constructions are discussed.

2. The Subjectivity Continuum. At the base of the analysis presented in this paper is the view that language is a subjective enterprise.5 All utterances are subjective to some degree in the sense that they presuppose a cogniser. But the degree of subjectivity is greater when the cognizer is encoded in the utterance, as in (2a).

(2) a. *It seems to me* that the weather is cold today.
    b. It's cold today.

Even in those cases where there is no overt cognizer as in (2b), a cognizing agent is implied. The difference is in the degree of subjectivity: (2a) is more subjective than (2b). In any use of language, there must be a speaker/cognizer, whose role is to provide the consciousness through which a pure 'objective' event is perceived and understood. The degree to which an utterance affords direct access to this consciousness--that is the degree to which it creates the impression that the contents of that consciousness are directly exposed--is what is meant by the phrase 'degree of subjectivity'. Thus the objective phenomena that occur in the world are liable to be expressed in a number of ways, each different expression coloring the event for a different degree of subjectivity.

    One might view events in the real world in terms of a typology of event phenomena. 'Action' is one type of real-world event; 'perceptual state', 'speech event' and 'emotional/psychological state' are others.6 These objective events may be expressed in different forms, varying in the degree to which a subjective viewpoint is encoded. The description of an event may be relatively external, or 'objectified,' as in *'John kicked the*

dog ;' or it may be relatively more internal, as in *'It looked to me as though John kicked the dog* .' In the first sentence the cognizer is placed outside of the event as an observer; on the other hand, the second type of sentence directly encodes the cognizer as a part of the utterance; the expression names the interpreter of the event. Schematically, this difference can be represented:

cognizer----> | X Y | | cognizer (=X') |
                                               | Y' |

less subjective <--------------------> more subjective

The difference between the agentive and experiential constructions lies in the degree of subjectivity. The agentive construction has an external observer/cognizer 'objectively' describing the psychological state of some experiencer (Suni in (1a)), whereas the experiential construction identifies the cognizer with the experiencer (Suni in (1b)), and is thus more subjective.

Although the encoding of subjectivity may vary from language to language, all languages seem to have the capacity to express differences in subjectivity. From the examples given in the preceding paragraphs it can be seen that English relies on syntactic periphrasis to express such differences. In contrast, Korean experiential and agentive constructions are examples of morpho-syntactic encoding of differences in subjectivity.

3. Morpho-Syntax. Table 1 gives a morpho-syntactic breakdown of typical predicates in the two constructions. It is important to note that for each predicate in the experiential column there is a lexically related predicate on the agentive side and that the morpho-syntactic relation is complex but systematic. In other words, Korean presents a speaker with a systematic set of coding options based on the same lexical material. In (A) -ha- is added to derive the agentive version from the experiential one; in (B) the experiential predicates are derived from the agentive ones with -i- or -ci- ; and in (C) both experiential and agentive predicates are derived from nouns.

	Experiential	Agentive	
	Adj-ta	**Adj-ha-ta**	
A.	kulip-ta	kulipwe-ha-ta	"miss, long for"
	musep-ta	musewe-ha-ta	"be afraid of"
	pulep-ta	pulewe-ha-ta	"be envious of"
	kwiyep-ta	kwiyewe-ha-ta	"love"
	silh-ta	silhe-ha-ta	"dislike"
	mukep-ta	mukewe-ha-ta	"feel heavy"
	tulyep-ta	tulyewe-ha-ta	"dread"
	elyep-ta	elywe-ha-ta	"be difficult for"
	mip-ta	miwe-ha-ta	"hate"
	yeppu-ta	yeppe-ha-ta	"be pretty, good"
	Noun+ha-ta	**Noun+ha-ha-ta**	
	pulanha-ta	pulanhae-ha-ta	"feel unsecure"
	yukweha-ta	yukwehae-ha-ta	"feel pleasant"
	ciluha-ta	ciluhae-ha-ta	"feel bored"
	pulkweha-ta	pulkwehae-ha-ta	"feel upset"
	sinkiha-ta	sinkihae-ha-ta	"be amazed"
	pilyoha-ta	pilyolo-ha-ta	"need"

Noun+iss-ta	Noun+iss-ha-ta	
masiss-ta	masisse-ha-ta	"feel tasty"
mesiss-ta	mesisse-ha-ta	"feel dandy"
caemiiss-ta	caemiisse-ha-ta	"be interested in"
B. **Verb-i-ta**	**Verb-ta**	
po-i-ta	po-ta	"see"
tul-i-ta	tut-ta	"hear"
ttel-i-ta	ttel-ta	"tremble"
Verb-ci-ta	**Verb-ta**	
nwiuchie-ci-ta	nwiuchi-ta	"repent"
nukkie-ci-ta	nukki-ta	"feel"
mite-ci-ta	mit-ta	"believe"
heyalie-ci-ta	heyali-ta	"figure out"
pese-ci-ta	pes-ta	"take off"
mancie-ci-ta	manci-ta	"touch"
C. **Noun-toi-ta**	**Noun-ha-ta**	
huhwe-toi-ta	huhwe-ha-ta	"regret"
kekceng-toi-ta	kekceng-ha-ta	"be concerned"
ihae-toi-ta	ihae-ha-ta	"understand"
hontong-toi-ta	hontong-ha-ta	"be confused about"
uysim-toi-ta	uysim-ha-ta	"be doubtful of"
sangsang-toi-ta	sangsang-ha-ta	"imagine"
Noun-sulup-ta	**Noun-ha-ta**	
salang-sulup-ta	salang-ha-ta	"love"
calang-sulup-ta	calang-ha-ta	"be proud of"
conkyeng-sulup-ta	conkyeng-ha-ta	"be respectful"

Table 1: Morpho-syntactic relations between experiential and agentive predicates

Table 2 gives the range of psychological predicates that enter into these parallel constructions. There does not seem to be any limitation, so that the constructions must be considered highly productive. Semantically the experiential construction extends into spatial metaphor (cf section 6). Morpho-syntactically the two constructions are exploited far beyond the domain of psych-verbs.

cognitive activity: thought, feeling, memory, etc	memory: *kieknata vs. kiekhata* thought: *saengkanata vs. saengkakhata*
cognitive disposition: easy, difficult, heavy, possible, etc	heavy: *mukepta vs. mukewehata* difficult: *elyepta v. elyewehata*
perception: see, hear, feel, touch, taste, etc	see: *poita vs. pota* hear: *tulita vs. tutta*
affective states: fear, pleasure, shame, boredom, etc.	fear: *musepta vs. musewehata* boredom: *ciluhata vs. ciluhaehata*
bodily sensation: cold, hot, hungry, sick, etc.	cold: *chupta vs. chuwehata* sick: *aputa vs. apahaehata*

Table 2: Semantic range of the parallel constructions

4. Interactional Discourse. Example (3) below is a series of potential turns in a conversa-tion between a speaker (A) and a listener (B) about a third person (Changho).

(3) A: *Na-nun ne-wa Changho-ka kathi issnun kes-ul poass-nuntey...* "I saw you with
   I-TOP you-and Changho-NOM together being-ACC saw-and...   Changho and... "

    a. *Na-nun Chango-lul cohahae.*   "I like him. " (AGT)
    b. *Na-nun Changho-ka coha.*   "I like him."(EXP)
    c. *??Ne-nun Changho-ka coha.*   "You like him. "(EXP)
    d. *Ne-nun Changho-ka coha?*   "Do you like him? "(EXP)
    e. *?? Changho-nun ne-ka coha.*   "Does he like you?"(EXP)
    f. *Changho-nun ne-lul coahani?*   "Does he like you?"(AGT)

It is well-known that the experiential construction is limited to use with first person subjects (cf. Kuroda 1973).[7] Turns a and b show that the speaker, A, can use either construction to describe his/her own feelings for Changho; (c) shows that the speaker A can not use the experiential construction to describe B's feelings about Changho, except if A is making an inquiry about B's feelings, as in (d), in which case the experiential construction is fine. But if the inquiry is about Changho's feelings for B, as in (e), the experiential construction is odd. Only the agentive construction can be used, as in (f).

    What unifies these facts is accessibility to the mental world of the experiencer by a cognizer, namely, the speaker. A, the speaker, can make 'direct access' assertions about his/her own feelings, or s/he can make an inquiry into the feelings of the interlocutor B, in effect making a request for a "direct access" conduit to B's mental state. Example (e) is odd in the experiential construction because neither A nor B have direct access to Changho's feelings. Examples (g) and (h) show that the experiential construction can be licensed by the presence of an evidential which specifies the conduit through which the speaker has access to the experiencer's state.

(3) g. *Ne-nun Changho-ka cohun ka poa*   "You like him, it seems."(EXP)
    h. *Changho-nun ne-ka coh tey.*   "He said that he likes you."(EXP)
    i. *Changho-wa na-nun ne-lul cohahae.*   "He and I like you."(AGT)
    j. *Changho-wa na-nun ne-ka coha.*   "He and I like you."(EXP)
    k. *?? Changho-wa na-nun kakca ne-ka coha.* "He and I each like you."(EXP)

In (g) this conduit is B's appearance and behavior; in (h) it is Changho's own words. What is asserted in (g) and (h) is the evidential link, not the experiential state itself. Thus (g) and (h) help to confirm that the basic issue underlying the choice between the two constructions is the accessibility of the experiencer's mental world to a cognizer, in this case the speaker.

    Examples (i) and (j) further clarify this point. In (i) A and Changho can be coded as a conjoined subject in the agentive construction with no implication about their personal relationship. But in (j), the experiential construction presupposes that A and Changho have intimate shared knowledge about B; for example, if A and Changho are siblings and B is the sister-in-law. This is confirmed by (k), where a distributive quantifier blocks the shared knowledge interpretation, making the experiential construction unacceptable.

5. Narrative Discourse. The facts of interactional discourse suggest a single unifying pragmatic factor controlling the occurrence of the experiential construction: accessibility of the experiencer's mental world to a cognizer. This factor is equally manifested in the use of the agentive and experiential constructions in narrative discourse. The 'accessibility' explanation offered for the limited distribution of the experiential construction in interactional discourse suggests that this construction ought to be used in narrative for a direct representation of the thoughts and perceptions of a character; in fact, it ought to be limited to such contexts. In contrast, the agentive construction ought to give an objective, external perspective on the character, even when presenting his mental state.[8]

It is a widely held view that a narrative text invites the reader to identify with a perspec-tive from which the story is presented.9 Narrative is made up of contexts which may be viewed as more or less subjective. So called objective context may be understood to be portrayed or reported by a narrator as an observer. Hence, the reader, adopting the narrator's point of view, also views the events of the story world as an observer. In subjective contexts, the reader can directly 'participate' in the story *via* the character's thoughts and perceptions. The character whose conscious-ness is adopted is called the 'subjective character,' and the phenomena that the reader experiences through that character are labeled 'represented thought' and 'represented perception' (Banfield 1982). Typically the text will present perceptions, thoughts and feelings without referring to the experiencer at all. The reader's focus is not on the experiencer as an actor in the story world but rather on the direct experience of thoughts, feelings and perceptions through that character.

We predict that the experiential construction will be used in subjective contexts for the straightforward expression of a subjective character's mental state, since in subjective contexts the reader has direct access to the subjective character's psychological state. We also predict that the agentive construction can occur in either objective contexts, or in subjective contexts to express a non-subjective character's mental state.10 The agentive construction is licensed in objective contexts because the reader, taking the narrator's perspective, views events as an external observer. The agentive construction in subjective contexts is possible to describe a non-subjective character's mental state because the subjective character cannot have direct access to any character's mental world except his own. (see ex 13 line 3)

Examples (4) and (5) are parallel episodes of narrative discourse. In (4) several linguistic features suggest represented thought of the subjective character *Suni*: the deictic *come...stand*, the reflexive *caki*, the non-embeddable fragment and exclamation, and the causal conjunction. In this context the experiential construction (a) is more coherent than the agentive (b) since it supports a constant psychological viewpoint from within Suni. In example (5) linguistic features such as the deictic *go...stand*, the adverbial *obviously*, the plain pronominal possessor and the progressive aspect all suggest an external, objectified view of Suni, and in this case the agentive construction (b) is coherent. This illustrates the basic effect of represented perception and thought versus objectified expression. The experiential contruction supports the subjective context because it conveys more direct access to the character's mental world, precisely the point of represented thought. On the other hand, the agentive construction supports the objective context, since it suggests an external view of the character.

(4) *Suni-nun kewul ap-ey wase sessta.*
   Suni-TOP mirror front-to come-stand

  papo kathi ulkin! ulkoissnun (caki) elkul-ul po-ni,
  fool like crying crying (self) face-ACC see-cause,

    a. *(Suni-nun) Toli-ka teuk miwessta.* (EXP)
      Suni-TOP Toli-NOM more hateful

    b. ?*Suni-nun Toli-lul teuk miwehaessta.* (AGT)
      Suni-TOP Toli-ACC more hate

  'Suni came and stood in front of the mirror. Crying like a fool!
  (Suni) seeing herself crying like that, Toli was even more hateful (to her) than before.'

(5) *Suni-nun   kewul ap-ey kase sessta.   kunye-nun punmyenghi ulko issessta.*
Suni-TOP mirror front-to go stand   she-TOP   obviously   crying was

ulkoissnun (kunye-uy) elkul-ul po-nmye,
crying   (her)   face-ACC see-and,

    a. *?Suni-nun Toli-ka miwessta.* (EXP)
    Suni-TOP   Toli-NOM more hateful

    b. *Suni-nun Toli-lul miwehaessta.* (AGT)
    Suni-TOP Toli-ACC more hate

'Suni went and stood in front of the mirror. It was clear that she
was crying. As she looked at her face crying, Suni hated Toli even more.'

      Examples (6 a, b and c) provide further evidence of the direct access function of
the experiential construction. The introductory sentence of (6a) sets up a represented
thought context for Insu in the following sentence. Here the experiential construction is
fully acceptable, just as it was in the represented thought context of example (4). Example
(6b) is projected thought; more specifically, it is Suni's report of Insu's mental state, as
Insu himself views it. Here the experiential construction is strange, but still marginally
acceptable. Finally, (6c) presents Suni's psychological report about Insu's mental state,
i.e. from Suni's point of view. Here there is no sense of direct access to Insu's mental
world and the experiential construction is impossible. Thus in moving from (6a) to (6c)
the context portrays decreasing access to Insu's mental world, and the experiential
construction becomes less and less acceptable.

(6) a. *Insu-nun ancase cangmi kkoch-ul pomye saengkakhaessta.*
   Insu-TOP sitting roses -ACC   seeing thought

     { *caki-nun i cangmi kkoch-ul coahanta.*(AGT) }
     { *caki-nun i cangmi kkoch-i cohta.*(EXP) }
     self-TOP this rose-ACC/NOM   like

    'Insu was thinking as he sat looking at the roses. (To him) These roses were nice.'

  b. *Suni-nun Insu-lopute manun iyaki-lul tuless-nuntey,*
   Suni-TOP Insu-from   many story-ACC heard-and

     { *ku-nun cangmi kkoch-ul coahanta.*(AGT) }
     { *? ku-nun cangmi kkoch-i cohta.*(EXP) }
     he-TOP   rose-ACC/?NOM   like

    'Suni heard many stories from Insu, you know, he liked roses.'

  c. *Suni-nun Insu-eytaehae manun iyaki-lul tuless-nuntey*
   Suni-TOP Insu-about   many story-ACC heard-and.

     { *ku-nun cangmi kkoch-ul coahanta.*(AGT) }
     { * *ku-nun cangmi kkoch-i cohta.*(EXP) }
     he-TOP   rose-ACC/*NOM   like

    'Suni heard a lot of stories about Insu, and he liked roses.'

The next piece of evidence has to do with aspectual inferences derived from the two constructions. In (7) the initial sentence sets up a character description. In the sequel only the agentive construction in (7a) is coherent. In contrast, the initial sentence in (8) sets up a narrative episode with a here and now deicticly anchored to the story world. In the sequel both constructions are coherent, but each has a different interpretation. (8a) is a psychological report giving a general disposition of Insu, explaining why he stopped, perhaps from another character's or the narrator's perspective. (8b), on the other hand, transparently presents what Insu is thinking at that moment of the story.

(7) *Insu-nun uskinun namca-ta.*
Insu-TOP funny   guy

    a. *Ku-nun cangmi kkoch-ul cohahaessta.* (AGT)
    b. *?? Ku-nun cangmi kkoch-i cohassta.*(EXP)
       he-TOP roses-ACC/?NOM like

    'Insu was a funny guy. He liked roses.'

(8) *Insu-nun kkoch cip   ap-ey   sessta.*
Insu-TOP flowershop front-LOC stood.

    a. *Ku-nun cangmi kkoch-ul cohahaessta.* (AGT)
    b. *Ku-nun cangmi kkoch-i cohassta.*(EXP)
       he-TOP roses-ACC/NOM like

    'Insu stopped in front of the florist to look. He liked the roses.'

The Agentive construction can describe a general psychological disposition, or it can summarize. As used in (7a), it gives a report about the mental state of the character, rather than tying us to the character's 'here' and 'now.' The lack of connection to a deictic center, a 'here' and 'now,' results in interpretations which are habitual, durative, timeless, or generic. On the other hand, the experiential construction, in directly representing a character's consciousness, is tied to that character's 'here' and 'now' in the story world. This leads to a punctual, perfective interpretation, as in (8b).

    Further support for the cognitive analysis of these constructions is provided by the use of the so called long-distance reflexive *caki* in the complement of psych-verbs, as in (9a and b). The reflexive *caki* expresses reflective consciousness in a context where the experiencer is an objectified element of his own consciousness (cf. reflective *Adam considered himself to be inadequate* versus non-reflective *Adam felt inadequate.*).11

(9) a. *Insu-nun* {*caki-ka / ?∅* } *Suni-lul silhehanta ko nnukkiessta.* AGT)
    Insu-TOP {self-NOM / ?∅ }   Suni-ACC dislike   that felt
    'Insu felt that he disliked Suni.'

    b. *Insu-nun*   {*?caki-ka / ∅* } *Suni-ka silhta   ko nnukkiessta.* (EXP)
    Insu-TOP {? self-NOM /∅ } Suni-NOM dislikable that felt
    'Insu felt Suni to be dislikable.'

In (9a) the agentive construction presents an objectified view of Insu's dislike for *Suni,* but since the whole is located within *Insu*'s mental world, as indicated by the matrix psych-verb, it is *Insu*'s objectification of his own feeling; he as an observer is regarding a separate intensional entity (experiencer), which happens to be himself. This reflective consciousness promotes the use of *caki* versus *zero* in the complement. On the other hand, in (9b) the experiential construction invites the reader to directly access Insu's

mental state. *Insu* himself is transparent, and *zero* is consequently more acceptable in the complement than is *caki*.

It is the usual case that each experiencer has complete access to his or her own mental world; hence, the preference for the experiential construction with first person subjects (see Section 2). An experiencer is less likely to have access to other human minds, but the potential for access does exist, especially in narrative. However, as Wierzbicka (1980) points out, we have little access to the minds of animals. The inaccessibility of animal minds is reflected in a comparison of (10) and (11).

(10) *Suni-nun twis kelum-ul chessta.*
    Suni-TOP back step-ACC took   'Suni stepped back.

    a. *Kunye-nun Insu-lul musewehassta.*(AGT)
       she - TOP Insu-ACC afraid-of   a. She was afraid of Insu.'
    b. *Kunye-eykey/nun Insu-ka musewessta.*(EXP)
       she - DAT/TOP Insu-NOM frightening   b. Insu was frightening to her.'

(11) *Ku kae-nun twis kelum-ul chessta.*
    The dog-TOP back step-ACC took   'The dog stepped back.

    a. *Ku kae-nun Insu-lul musewehassta.*(AGT)
       the dog-TOP Insu-ACC afraid-of   a. It was afraid of Insu.'
    b. ?*Ku kae-eykey/nun Insu-ka musewessta.*(EXP)
       the dog-DAT/TOP Insu-NOM frightening   b. Insu was frightening to it.'

(10) shows both constructions are acceptable with a human experiencer if a context for represented thought is set up. (11) shows that in the same context the experiential construction (11b) is odd with an animal experiencer, unless, of course, the whole is embedded in a story such as London's "Call of the Wild," which personifies the animal character.

The following data, taken from Korean narrative stories, instantiate the claims that have been made in this paper. These examples provide some illustration of the narrative effects available through the use of experiential and agentive constructions. (12) is taken from an episode in which the reader participates in *Yuceng*'s fear of *Molan*.

(12) 1.*Yuceng-un enusai mom-ul tosalinta. sulmyesi kongpo-ka moliewassta.*
        Yuceng-TOP abruptly body-ACC withdrew. slowly fear-NOM crowd-came

    2. [*nunap-ey anca cangnankkuleliki-chelem cocaltaenun*]
       eye-front-LOC sitting pranster-like   chatting

    3. *Molan-i kapcaki museweciessta.* (EXP)
       Molan-NOM suddenly fearsome

    '1. Yuceng abruptly withdrew. Slowly fear came crowding in.
    2. There sat (Molan), chattering like a mischevious kid.
    3. Suddenly Molan was frightening (to her).'

The context is represented thought, with Yuceng as the source. In line (1) the spatial metaphor (see Section 6) *fear came crowding in* helps the reader to identify with her emotional state. In line (3) the experiential construction with deleted experiencer continues the identification with Yuceng as a subjective character by directly representing her thought process.

Example (13) is necessarily a bit more complex due to the contrast between the two construction types. The entire passage is represented thought, with 'he' (Youngha) as the subjective character. Youngha is thinking about Yuwha and Yuwha's hatred toward Hyenwu. Line (1) contains the experiential construction, which directly represents Youngha's mental state: Youngha's concern for Yuwha. In line (3), the agentive construction gives Youngha's projection of Yuwha's mental state: Yuwha hates Hyenwu. The reader has direct access to Youngha's mind, hence the experiential construction, but not to Yuwha's mind, hence the agentive construction.

(13) 1. **ku-*nun* *Yuwha-ka* *kekcengsulewessta.*** (EXP)
      he-TOP Yuwha-NOM concerned. (He = Youngha)

2. [*Kewuy kwangki-ey kakkawul*] *cengto-lo salanamun*
   almost madness-to near   degree-with survived

3. *Hyenwu-lul cungohakoiss-*nuntey, (AGT)
   Hyenwu-ACC hating-but

4. [[*salanghanun salam-ul ilun*] *sulpum ttaemun-ila*] *ko haki eynun*
     loving   person-ACC lost sadness because-is   to-say

5. *com   cinachita siphessta.*
   somewhat above   seemed.

'(1) Yuwha was a concern to him [Youngha] (3) Although (Yuwha) was hating Hyenwu, (2) the survivor, with near madness, (5) it [her hatred] seemed (to him) to be more (4) than the sadness of losing her lover.'   (No Won p.46)

6. Cognitive Models. The analysis given in this paper follows Gerdts and Youn (1988) in positing distinct underlying structures for the two constructions. Their unaccusative analysis of the experiential construction contains a locative case-marked experiencer in the initial stratum, while the corresponding agentive construction portrays the experiencer as an agent. This initial-level case marking suggests distinct cognitive models for the two constructions. Table 3 provides a simple schematic of what we think the folk cognitive models underlying the two constructions might look like.[12] Specifically, the experiential construction has a locative basis, while the agentive construction has a force-dynamic one. In the experiential construction the experiencer is categorized as a location which is approached by a stimulus, the object of experience. In some cases this stimulus emanates from a source.

	experiencer	object of experience
Experiential construction	*motion* LOC <----------	Stimulus (Source)
Agentive construction	*energy* AGENT --------->	Object

Table3: Folk models of mental events

That Korean experiencers metaphorically conceive of themselves as locations approached by stimuli is shown by the extensive use of "*ota*" (to come) and "*tulta*" (to

enter) in predications of mental events, as in (14),[13] which follow the distribution of the experiential construction.

(14) a. *sulmyesi kongpho-ka (Suni-eykey) molie-wassta.*
     slowly   fear-NOM (Suni-DAT) gather-come
     'Slowly fear gathered in to Suni.'

  b *nolae soli-ka*   *(Suni-eykey) tulie-wassta.*
    song sound-NOM (Suni-DAT) heard-come
    'Suni heard the sound of music coming to her.'

  c. *kulen nukkim-i*   *(Suni-eykey) tulessta.*
    that feeling-NOM (Suni-DAT) enter
    'That kind of feeling came to Suni.'

  d. *mom-i*   *(Suni-eykey) ttelie-wassta.*
    body-NOM (Suni-DAT) shiver-come
    'shiver overcame Suni.'

  e. *(Suyen-eykey) Min paksa-uy mosup-i tteola-wassta.*
    (Suyen-DAT) Min Dr.-GEN figure-NOM surface-come
    'The figure of Dr. Min came to Suyen ('s mind).'

  f. *kapet-uy*   *kamkak-i*   *(Kanguk-eykey) cenhae-wassta.*
    carpet-GEN touching-NOM (Kanguk-DAT) transmit-come
    'Kanguk became aware of how the carpet felt.'

  g. *papokathun saengkak-i (Suni-eykey) tulessta.*
    foolish   thought-NOM (Suni-DAT) enter
    'A foolish thought came to Suni.'

Furthermore, the case-marking pattern of the experiential construction as in (15) is exactly parallel to that of a locative expression, even when it is not intuitively obvious that a stimulus is approaching a location. Compare the literal locative in (15a), and the experiential construction in (15b).

(15) a. *Toli-eykey sopho-ka*   *wassta.*   b. *Toli-eykey Suni-ka*   *musepta.*(EXP)
    Toli-DAT package-NOM come   Toli-DAT Suni-NOM fearful
    'A package arrived to Toli.'   'Suni is fearsome to Toli.'

     An interesting phenomenon to consider in this respect is the fact that body part metonymy (Lakoff 1987) is not only possible, but highly frequent in the experiential construction. The body part used in the expression is the relevant organ to which the stimulus must come in order for it to be experienced. This model for perception is depicted in Table 4.

Type of Experience	body part metonymy	example for experiencer
sight	eyes	*pulpich-i nun-ey pointa.* light-NOM eye-DAT visible 'The light was visible to his eyes.'
sound	ears	*phili soli-ka kwi-ey tulinta.* recorder sound-NOM ear-DAT heard 'The sound of a recorder was audible to (his) ear.'
touch	skin	*chukchukhan kamkak-i phipu-ey nukkiecinta.* wet feeling-NOM skin-DAT felt 'Wet feeling was sensible to (his) skin.'
affect	heart/mind	*Suni-ka kasumsok-ey kulipta.* Suni-NOM heart-DAT longed-for 'In his heart was longing for Suni.'
thought	head/brain	*Suni-uy saengkak-i meli-ey nassta.* Suni-GEN thought-NOM head-DAT strike 'The thought about Suni struck him.'

Table 4: Body part metonymy for experiencer

Instead of having the experiencer specified in the dative/locative, a visual experience can come to the eyes, an auditory experience to the ears, a tactile experience to the skin, affect to the heart or body, and thought to the head or brain. This systematic metonymy emphasizes the experiencer as a location since the actual site of experience, in the folk model, is substituted for the whole person, who is usually not mentioned. (Note the lack of overt possessor in Table 4). In contrast, this body-part metonymy is not possible in the Agentive construction. This reflects the folk assumption that people, and not any specific part of them, are the default source of agency. Note the weirdness in English of '*his mind wanted to leave*,' or '*his foot kicked open the door*'.

Finally, the case marking of the experiential construction places focus on the object of experience, and not only backgrounds the experiencer in an oblique expression, but also promotes its deletion in both interactional and narrative discourse. This is consonant with the claim that the experiential construction allows the reader to view the story world through the thoughts and perceptions of the experiencer, who remains more or less transparent. In constrast, the agentive construction portrays the experiencer as an agent, i.e. as a salient actor in an objective presentation of the story world.

7. Conclusion. In this paper we have attempted to account for two types of constructions in Korean. The difference between experiential and agentive constructions is signalled at the surface level by systematic differences in derivational morphology and case marking. The semantic distinction is accounted for in terms of the notion of subjectivity: the agentive construction has an external cognizing observer who has no access to the mental world of the experiencer, whereas the experiential construction incorporates the cognizer as the experiencer in the utterance who has direct access to her own mental state. We also have discussed the effect of these semantic and pragmatic differences in interactional and narrative discourse. The experiential construction is exploited in subjective contexts in narratives to give the reader the feeling of direct participation in the story world, whereas the agentive construction is used in objective

contexts where the reader is acting as a mere observer of the scene just like a narrator. A locative/directional model for the experiential construction is proposed, and a force dynamic model for the agentive construcion. One avenue for future research is an investigation into the similarity of the case pattern [DAT-NOM-PREDICATE] used both in the experiential construction discussed here and in passive constructions, and on the stativity vs. activity of these constructions.

NOTES

1 The syntactic constructions in question are not limited to psych-verbs. In particular the structure in (1b) with overt action predicates is described as "passive." See also the discussion of Table 2.
2 Where possible we translate the experiential construction in Korean with a superficially parallel inversion construction in English. This does not, however, capture either the distributional possibilities or the semantic effects of the Korean, which are central topics of this paper.
3 Experiential constructions have been noted in Russian, Japanese, Georgian and German, in addition to Korean. They have variously been called Psych Constructions, Inversion Constructions, Double-Subject Constructions, Dative Subject Constructions, Impersonal Constructions or Unaccusative Constructions. See Postal 1970, Youn 1985, N. McCawley 1975 for examples of previous treatments of this phenomenon.
4 N. McCawley's feature "unself-controllable/non-volitional" is one of the more interesting proposals. Lack of volition is a property of the experiential construction which we do not explore in this paper. But it seems easily derivable from the cognitive semantic basis of the construction which we propose in section 6. In particular (cf Table 3), the force-dynamic basis of the agentive construction portrays the experiencer as the controller of the event. In contrast, the locative basis of the experiential construction, and the locative case-marking on the experiencer, seem to invite the inference that the experiencer is a mere location of the experience, and by contrast with the agentive construction, not in control of it.
5 The term 'subjective' is naturally understood in relation to its opposite notion 'objective,' and while these contrasting notions may be useful in understanding the cognitive models behind the experiential and agentive constructions, this should not be taken to imply that language itself can be objective.
6 There may be other categories; these are offered only for purposes of illustration.
7 German shows a similar restriction, but only with a few archaic predicates. For example in interactional discourse, *mich dünkt* is acceptable, but **dich dünkt* or **ihn dünkt* are not possible, although in narrative *ihn dünkte* suggests direct access to the mental state of the character. In contrast, the verb *denken*, with an 'agentive' case-marking pattern, has no such restriction..
8 This point is essentially parallel to Kuroda's claim about adjectival versus verbal predicates in Japanese narrative.
9 This view is implicit in Fillmore's (1975) work on deixis; it is also articulated by Banfield (1982) and Cohn (1978) among others in literary criticism and by Kuno(1987), Kuroda(1973) and others in linguistics.
10 The agentive construction appears to occur in subjective contexts to describe the subjective character's mental state. But rather than directly representing the thought or perception of the character, it gives a psychological report (Wiebe, 1989) of his general

psychological disposition from an external 'objective' point of view (cf examples 7a & 8a). The agentive construction also occurs in passages of **reflective thought,** in which the subjective character is the observer of his own objectivized mental state, as in example (9a).

11 See Zubin, Chun and Li (1990) in this volume for a fuller discussion of this point.

12 We are indebted to pioneers in Cognitive Semantics such as Ronald Langacker and Leonard Talmy for their insights into the underlying cognitive structure of these constructions, and in particular for insisting that such constructions are not maps of reality, but rather conventionalized *construals* of it.

13 See Li, in preparation, for a perspective taking account of experiential constructions in Mandarin.

REFERENCES

BANFIELD, ANN. 1982. Unspeakable Sentences. Boston: Routledge and Kegan Paul.

COHN, DORRIT. 1978. Transparent Minds. Princeton: Princeton University Press.

FILLMORE, CHARLES. 1975. Santa Cruz Lectures on Deixis: Bloomington: Indiana Linguistics Club.

GERDTS, DONNA B. and CHEONG YOUN. 1988. Korean Psych constructions: Advancement or Retreat. CLS 24.155-75

KURODA,S.-Y.1973. Where Epistemology, Style and Grammar Meet: A Case Study from Japanese. in S. Anderson and P. Kiparsky, eds. A Festschrift for Morris Halle. 377-91.

KUNO, SUSUMU. 1973. The Structure of the Japanese Language. Cambridge: MIT Press.

KUNO, SUSUMU. 1987. Functional Syntax:. Chicago: U. of Chicago Press.

LAKOFF, GEORGE & MARK JOHNSON. 1980. Metaphors We Live By. Chicago: University of Chicago Press.

LAKOFF, GEORGE. 1987. Women, Fire and Dangerous Things. Chicago: U. of Chicago Press.

LANGACKER, RONALD. 1985. Oberservations and Speculations on subjectivity. In Haiman, J. (ed). Iconiciy in Syntax:. Philadelphia: John Benjamins Publishing Co.

LI, NAICONG. in prep. "Perspective taking in Mandarin narrative". Ph.D. dissertation, University at Buffalo.

McCAWLEY, NORIKO A. 1975. What strikes me about Psych-movement. in P. Reich (ed.) LACUS 2. 320-28.

POSTAL, PAUL. 1970. On the Surface Verb 'Remind'. Linguistic Inquiry 1.

RAPAPORT, WILLIAM J., et al. 1989. Deictic Centers and the Cognitive Sturcture of Narrative Comprehension: Tech Report 89-01: Dept. of Computer Science, University at Buffalo.

WIEBE, J. 1989. Recognizing subjective sentences.Unpublished PhD. dissertation, University at Buffalo.

WIERZBICKA, ANNA. 1980. Lingua Mentalis. New York: Academic Press.

ZUBIN, DAVID A., SOON AE CHUN & NAICONG LI. 1990. Misbehaving Reflexives in Korean and Mandarin. BLS 16 (this volume).

TEXT REFERENCES

[Han Soosan] HAN, SOOSAN. 1980. Haepingki-uy achim. vol 2. Seoul.

[No Won] NO, WON. 1980. Paesin-uy Kyaycel. Seoul: Minyaysa.

Experimental Evidence Of The Transfer Of L1 Implicature In L2 Acquisition

Ron Cowan & Brad Reed
University of Illinois

This paper discusses what we believe to be an interesting example of a pragmatic constraint on the use of a specific syntactic structure. In order to accept our use of the term `pragmatic' one must be prepared to grant that in linguistics pragmatics is concerned with the planning of communication as well as its interpretation. We recognize that some readers may be unhappy with this, but we feel that the phenomenon we are describing belongs as much to pragmatics as it does to discourse, since, in one instance, it has a type of implicature which is neither completely conversational or conventional. The constraint we will describe is regulated by the speaker's sensitivity to specific contextual features in the ongoing discourse. Our modest goal will be to characterize the conditions which could lead to the development of the constraint. In the course of pursuing this goal we will first demonstrate the robustness of the constraint by showing that it is subject to transfer in second language learning. We will then examine changes associated with this structure which provide a basis for our hypothesis of how this constraint can arise. Finally, we will provide cross-linguistic evidence which supports our hypothesis.

The syntactic structure in question is the so-called *bei* passive in Mandarin. As 1 shows, it involves subject-object inversion with the concomitant insertion of the particle *bei*.

(1)　a.　Zhang　da　le　　　wo　　　　　`Zhang hit me'
　　　　　(name)　hit　perf.　me

　　　b.　Wo *bei*　　Zhang　da　le　　　`I was hit by Zhang'
　　　　　I　　pass.　(name)　hit　perf.

Like its English counterpart, the *bei* passive may delete the agent NP; however, since Mandarin is a topic prominent language, the *bei* passive is not used to highlight objects as the English passive often is. This can be accomplished by the topicalization transformation in 2, which is identical to the English 'yiddish movement' transformation, said to exist in all languages.

(2)　　　Zhang　wo　yi jing　kan　le.
　　　　　(name)　I　already　see　perf.

　　　　　`Zhang, I've already seen'

In subject prominent languages like English, the passive is often employed to bring an object into focus when it has been established as the theme or topic of discourse. Thus sentence 3a is generally preferred over 3b because it maintains the theme-rheme flow, or what is often referred to as the 'given-new contract' (Haviland & Clark 1974) in the preceding discourse. In Mandarin the *bei* passive could not be used for this purpose. Instead, some other device such as the *shi...da* construction in 4a or the construction shown in 4b would be employed.

(3)　　　It had been one lousy morning. Bill had gotten up
　　　　　late for work. Rushing through breakfast he had
　　　　　burned his toast twice. When he got to the office he
　　　　　found a nasty letter from the IRS, and then to top it all
　　　　　off...

a. He was bitten by my pet cobra, Coral.
b. Coral, my pet cobra, bit him.

(4)   Jien tien women zhai ke tang shan tao lun   yi   ben
        today   we   in   class   discuss   one   class.

        shu ta miao xei   le   zhang zheng   gei ren min
        book it describe   perf.   war   cause   people

        dai lai   de   ku nan
        bring   nom.   hardship

        `In class today we were talking about a book that
        describes the war. It relates the hardships people
        endured then'

        a. Nei ben   shu   shi wo mu gin xie   de
           that class. book   is my mother write   nom.

           `That book was written by my mother'

        Wo yuanlai dasuan xie   yi ben   yuanguan
        I originally planned write one class. about

        Tian'an men shijiande shu keshi ni zhidao ma
        Tianamen   incident book but you know ques.

        `I had originally planned to write a book about the
        Tienamen incident, but, you know...

        b. Nei ben   shu   yi jing cu ban   le
           that class. book   already published   perf.

        `That book has already been published'

The *bei* passive is subject to lexical constraints. The logical object tends to be animate (Cheng personal communication) and the verb normally describes how the object is `dealt with, manipulated or handled in some way' (Li & Thompson 1981:501). A sentence like 5 is thus ungrammatical because the verb does not express disposal. In addition to these lexical constraints the *bei* passive is subject to the constraint which is the topic of this paper — it occurs almost exclusively in discourse contexts that express adverse or unfortunate events (Chao 1968:72, Li & Thompson 1976:467, Li & Thompson 1981:493). Thus the Chinese equivalent of the English discourse in 3 could be appropriately completed with the *bei* passive sentence in 6 since it describes a series of unfortunate events. But it is not possible to use a *bei* passive to complete either discourse in 4, due to the absence of elements of adversity. Li & Thompson (1981: 495) point out that this constraint is so strong that a sentence like 7 with a neutral verb would, in the absence of a preceding discourse, have the implicature that the speaker did not want to be seen.

(5)   * Ta *bei* ta   de   tong xue   hen   le
         he pass. his   nom.   classmates   hate   perf.

         `He is hated by his classmates'

(6)	Ta *bei*	she	yao	le
	he pass.	snake	bit	perf.

'He was bitten by a snake'

(7)	Wo *bei*	lao shi	kan jian	le
	I pass.	teacher	see	perf.

'I was seen by the teacher'

An interesting question for theories of second language acquisition and discourse processing is whether this constraint associated with the *bei* passive would be transferred by Chinese speakers learning an SVO subject prominent language like English. Judging from examples 4 through 7, a reasonable hypothesis might be that the Chinese learners will be more sensitive to the situations described in an English discourse than to the topicality signals in deciding whether or not to use an English passive. To test this hypothesis, we ran an experiment with 36 native speakers of Chinese who were studying at the University of Illinois and 18 native speakers of English, also students, who served as a control group. The Chinese subjects were divided into two groups — intermediate (with a mean TOEFL score of 530) and advanced (with a mean TOEFL of 586) — on the basis of their English proficiency. All of the subjects viewed 60 sequences of three slides in Chinese and English, 40 of which were experimental sequences and 20 of which were distractor sequences. Experimental sequences were classified as positive or negative depending upon the contextual features of the episode depicted. Negative episodes were operationally defined as those in which the subject was in some way adversely affected or some negative action, such as a theft, took place. Positive episodes were defined as those in which there was no such adverse or negative effect or where the action was of a positive or neutral nature such as someone helping someone else.

For the first two frames of every sequence a tape recording in Chinese and English described the actions being portrayed. Below the scene depicted in the final slide were two sentences, one an English passive or a *bei* passive, the other, an active sentence in English or Chinese. A typical sequence, with English and Chinese recordings written under the first two slides, and a Chinese and English version of the third slide is shown in the Appendix.

Sequences were balanced to insure an equal number of animate and inanimate objects and situations where the topic was up or down. By topic up is meant that the passive sentence brought the object into focus so that it matched the theme of the preceding sentence as in 8. In topic down situations the logical object either did not match the theme or had not appeared in the preceding sentence.

(8) Example of a topic up sequence (Negative situation)
Slide 1: A boy was walking.
Slide 2: The boy didn't watch where he was going.
Slide 3: a. The boy was hit by a truck.
      b. A truck hit the boy.

The Chinese subjects were assigned to one of four groups composed of from seven to ten subjects. Each group received an English test running order and a Chinese test running order. The control group subjects were also assigned to four groups. Each group was given sixty-item response sheets. The subjects were told that they were participating in a test of style and that they were to indicate their preference for one sentence or the other in the third slide by marking either (a) or (b) on the response sheet. They were instructed to indicate their interpretation of each sequence of three slides as either negative or positive by marking a plus or a minus beside each item on the response sheet. Negative and positive situations were defined for the subjects in the manner

shown. The sequences were then projected while the accompanying sentences were played. A three to five second pause separated each frame and a seven second pause separated sequences. The Chinese subjects were given the Chinese test first and the English test seven days later. All subjects were paid for their participation.

Since the subjects' interpretations of the episodes as positive or negative were important to the study, the first thing we did was correlate their responses in this regard to the original sequence divisions on the target sequences. The correlations were high for both groups — .82 for the Chinese subjects and .88 for the American subjects. The first analysis of variance (ANOVA) was run on just the Chinese speakers' performance on both the English and Chinese test. The factors involved were language (Chinese vs. English), proficiency (intermediate vs. advanced) situation (negative vs. positive), object (animate vs. inanimate), topic (up vs. down) and sentence type (active vs. passive). There were significant main effects for situation $F(1,34) = 69.05$ p < .00001, object F $(1,34) = 6.4919$ p < .015, and topic $F(1,34) = 19.12$ p < .0001. The lack of a main effect for language simply indicates that the Chinese subjects did not perform any differently on the English or the Chinese sequences with regard to their selection of passives. Similarly, the lack of a main effect for proficiency indicates that there was no significant difference in performance on both tests attributable to knowledge of English. There were a number of very significant interaction effects, all of which looked like the one in Figure 1, which shows the Chinese subjects performance operating across languages and proficiency levels.
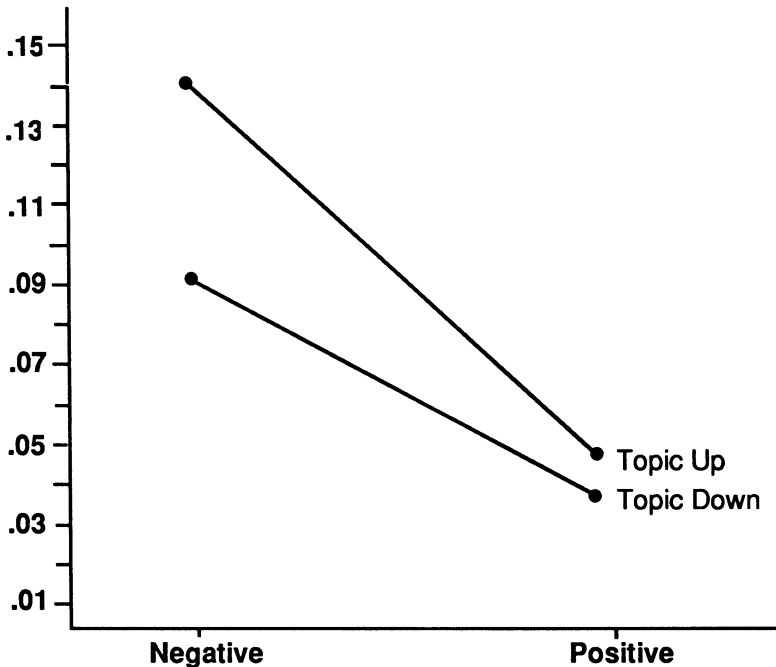
Figure 1
Situation x Topic Position Interaction

It is clear that situation is the dominant influence, with negative situations constituting the optimal condition for use of the passive. Within the negative situation, topic position exerts a stronger influence than it does in positive situation.

Of course this does not prove that Chinese transfer their preference for the passive in negative situations in second language learning. To discover if this is in fact the case, we must compare the Chinese speakers' performance on the English passive with that of the control group, which is what the next ANOVA did. Here we found an extremely strong main effect for language group $F(1,52) = 10.65$ p < .0019, which reflected the Chinese subjects' significantly higher choice of the passive than the English speakers'. However the decisive evidence is the interaction $F(1,52) = 5.273$ p < .025 between language group and situation shown in Figure 2. This clearly supports the claim that the Chinese speakers were transferring their preference for the connection between passives and negative situations to the second language. Although the American control group did choose the passive more often in negative than in positive situations, the difference is quite small, suggesting that the Americans are not sensitive to the situation as a whole.
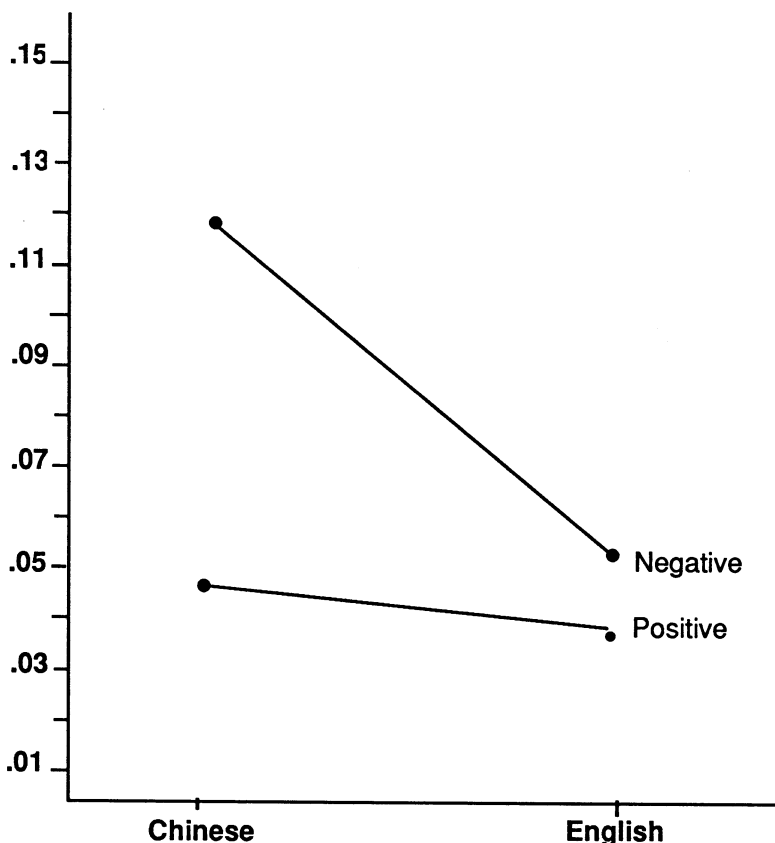
Figure 2
Language Group x Situation Interaction

To clarify this tendency among the American control group subjects, an ANOVA was performed on their data alone. The factors included situation, object, and topic position. Predictably, the only factor showing a significant effect alone was topic position $F(1,17) = 14.655$ p $< .0013$.

The results indicate that the Chinese subjects' preference for using the *bei* passive in discourse describing negative or adverse situations is robust and is transferred to the grammatical construction that is seen as formally equivalent in the the target language — the English passive. Even those Chinese subjects most proficient in English were more sensitive to the situation than the topicality signals in processing discourse in a second language. These results are important for discourse processing and second language learning because they constitute one possible cause for second language learners' inability to master the organizational structure of the target language discourse.

Given these findings, one might hypothesize that this constraint governing the use of the *bei* passive is fairly resistant to change. But, in fact, it appears that the *bei* passive is being gradually extended to non-adverse contexts on Mainland China as a result of translations from Russian and English documents. In another experiment which used the same procedure but only Chinese sentences (Reed & Cowan 1989), we found that Taiwanese subjects were more constrained in their choice of the *bei* passive in negative situations than in positive or neutral situations than Mainland Chinese subjects. Although Mainland Chinese subjects selected *bei* passives more often in non-adverse contexts than the Taiwanese subjects, the difference between the two groups was not significant. We were thus forced to conclude that we had only weak evidence for some change in the restriction governing the use of the *bei* passive on Mainland China as opposed to Taiwan. Had our experiment included agentless passives, we might have been able to obtain a more revealing result with regard to the two Chinese populations. Recently it has been noted (Pei & Chi 1987) that agentless *bei* passives are being extended to non-adverse contexts. Thus, for example, a verb like tell, which was formerly restricted to active constructions like that shown in 9a now regularly appears with the *bei* passive as in 9b.

(9)   a.   Renjia gaosugou wo   `People have told me'

      b.   Wo *bei* gaosu   `I have been told'

This development has its origin in the translations of English reports made available to the Chinese news agency by the international wire services. The translators who work for the widely broadcast Reference News, have, over time rendered agentless English passives like `it was reported...' in the Chinese *bei* passive. However this expanded use of agentless *bei* passives has been restricted (to date) largely to writing, and it remains to be seen whether it represents the first stage in the breakdown of the situational constraint which pertains to *bei* passives with agents.

This development brings us to our final point, an account of what leads to the constraint governing the use of the *bei* passive.[1] Our hypothesis is that when a language possesses more than one way of maintaining the given-new contract in discourse, and the passive is not the primary means employed for this purpose, the passive with agent, which is the marked form of this construction is likely to undergo the kind of development we have seen with the Chinese *bei* passive. This implies that the unmarked form, the agentless passive, represents a conscious choice by the speaker to de-emphasize the agent. This is often done when the speaker wishes to report something that is a generalization or statement about which no direct reference to the agent needs to be made.

Cross-linguistic evidence supporting our hypothesis is found in a number of South Asian languages. Pandharipande (1976, 1982) has pointed out that in Hindi, Marathi and Nepali the passive with agent has taken on a specific meaning of capability which is determined by agent internal conditions, such as headache, hatred, happiness, physical or

psychological pain. etc., that is, various states which would could result in a volitional act. Note that this capability interpretation of the passives with agent in examples 10a-c is not expressed in the corresponding active sentences in 10d-f.

(10)

Hindi: a. Mujh se kuch bhī kahā nahī gayā
                 me by anything said   not went

Marathi: b. Mājhyā   kadūn   kahīhī bolla gela nāhī
                 me   by   anything said went not

Nepali: c. Mabāta   kehi   bhaniena
                 me by anything   said+pass.+m. pl.+ not

               'I wasn't able to say anything'

Hindi: d. Mɛ   ne kuch bhī nahī   kahā
                 I   ag. anything   not   said

Marathi   e. Mi   kāhīhī   mhatla   nāhī
                 I   anything said   not

Nepali   f. Maile kehi   (pani) bhanīna
                 I ag.m.anything (too) say+pt.+neg.

               ' I did not say anything'
         * 'I was not able to say anything'

What makes this case virtually identical to the one we have described for the *bei* passive is that the passive with agent is blocked when the preceding context indicates that the agent does not have any control over the act expressed by the verb. In Hindi opening the door, which is normally considered a volitional act, is performed `unknowingly or under the influence of alcohol' in 11a. It must therefore be expressed in an active sentence since these contextual circumstances exclude the passive version shown in 11b. Agentless passives in these three languages will either be interpreted presciptively i.e. `one doesn't do that' or, if the appropriate context has been set up, as having the volitional reading.

(11)

  a. Anjāne mẽ   us ne darwāzā kholā ɔr vah
        naše kī dhundh mẽ
        'Unknowingly
        'Under the influence of alcohol   he opened the door

        bāhar calā gayā.
        and went out'

  b. ?? Anjāne mẽ   (us se) darwāzā kholā gayā ɔr
        naše kī dhundh mẽ
        vah bāhar calā gayā

        'The door was opened by him unknowingly/under the influence of alcohol'

Obviously we welcome data from other languages which would assist in refining this hypothesis that passives with agents are likely candidates for developing the syntactic-pragmatic intersection we have described.[2] For the present we prefer to frame the hypothesis in its weakest form, that is, as a possible development which can occur irrespective of language typology if the appropriate conditions are present. Confirmation of the hypothesis would provide at least a small increment in our understanding of the mechanisms of diachronic syntactic change, and it would demonstrate one way in which syntactic-pragmatic phenomena can override universal conditions governing the organization of discourse in the production of natural language.

Notes

1.   The authors wish to thank Hans Hock for his very insightful observations about passives in different languages. His comments greatly aided the development of the hypothesis proposed in this paper.

2.   Dan Slobin has indicated that there is additional evidence from Turkish that supports this hypothesis. Because of the deadline constraints we were unable to incorporate this evidence in our paper.

REFERENCES

Chao, Yen Ren. 1968. A grammar of spoken Chinese. Berkeley: University of
              California.
Haviland, Susan E., and Herbert H. Clark. 1974. What's new? Acquiring new
              information as a process in comprehension. Journal of Verbal Learning and
              Verbal Behavior 13.512-521.
Li, Charles N., and Sandra A. Thompson. 1976. Subject and Topic:: A new typology
of   language. Subject and topic, ed. by Charles N. Li, 457-89. New York:
              Academic Press.
_____.1981. Mandarin Chinese: A functional reference grammar. Berkeley:
              University of California.
Pandharipande, Rajeshwari. 1976. Passive as an optional rule in Hindi, Marathi, and
              Nepali. SALA 1, 89-106.
_____ .1982. Volitionality: More evidence for constraints on passive in Hindi.
              SALA 6. 80-103.
Pei, Zhou Zhi, and Feng Wen Chi. 1987. On nativization of English. World Englishes
              6.111-26.
Reed, Bradly, and J R. Cowan. 1989. On applying an experimental approach to the
              investigation of syntactic change. Georgetown Round Table on Languages
              and Linguistics 1988, ed. by Thomas J. Walsh, 260-67. Washington D.C.:
              Georgtown Univ. Press.

APPENDIX
Sample Negative Sequence

Sequence #3--negative
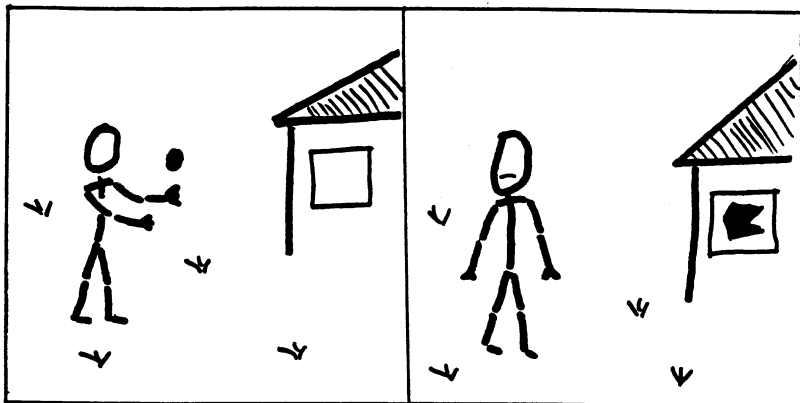
a) (A boy was playing with a ball.)

（一個小男孩在玩球）

b) (He broke a window with the ball.)

（他用球打破了玻璃）
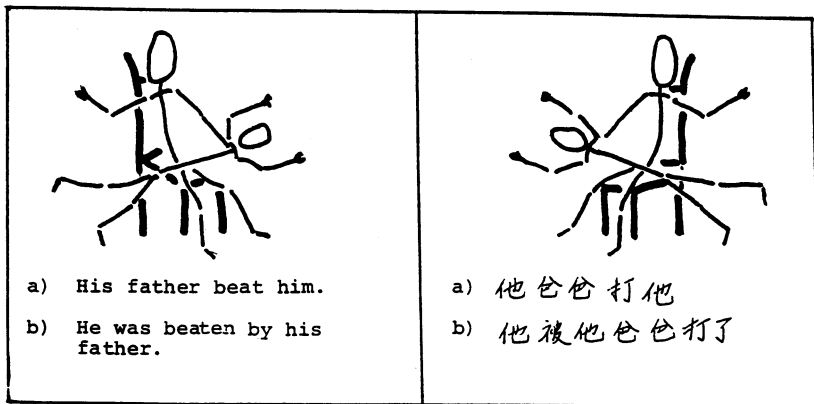
a) His father beat him.

b) He was beaten by his father.

c) English

a) 他爸爸打他

b) 他被他爸爸打了

c) Chinese

THE DIACHRONIC DEVELOPMENT OF SEMANTICS IN COPULAS

Dan Devitt
State University of New York at Buffalo

A well-known and much-discussed feature of Spanish is the use of two separate verbs- *ser* and *estar* - which are glossed as equivalents to the English 'to be'.[1] Reference grammars of Spanish traditionally describe the distinction between *ser* and *estar* in terms of a semantic difference that is implied in their use. *Ser* is used to imply a relatively permanent condition or state; *estar*, on the other hand, is used to imply that the condition or state predicated of the subject is temporary, as shown in Example 1.

(1) a. Juan es enfermo   'Juan is sick.' (implying Juan is sickly, an invalid)
   b. Juan esta enfermo   'Juan is sick.' (implying Juan is sick right now, but will recover eventually)

Turkish, like many languages, does not generally use a copula for present tense expressions of equivalence. A simple juxtaposition of subject and complement is the more frequent form of the expression, especially in the spoken language. In writing and formal speech, however, the copula may be included optionally. With third person subjects, the copula is the enclitic form-*dir,* which adds an epistemic modal flavor of probability to the sentence, as can be seen in Example 2.

(2) Ahmet iyi-dir   'Ahmet must be well; surely Ahmet is well.'

An interesting etymological parallel can be noted here: both Spanish *estar* and Turkish *-dir* can be traced back historically in their respective languages to verbs meaning 'to stand'.

The question that this paper addresses is whether this fact is nothing more than a coincidence, or an indication of some deeper parallel between the two. The claim put forward here will be that these two senses of the copula (i.e., the 'temporary' sense observed in connection with the Spanish copula *estar* and the 'presupposition' meaning that characterizes the use of the Turkish enclitic *-dir*) are not simply isolated uses of copulas in two unrelated languages. The position that I adopt is that these two uses can be related as different stages on the same evolutionary path of semantic development, illustrated in the figure below.

| POSTURAL VERBS | LOCATIVE VERBS | EXISTENTIAL VERBS | COPULA with a TEMPORARY SENSE | EPISTEMIC MODAL |
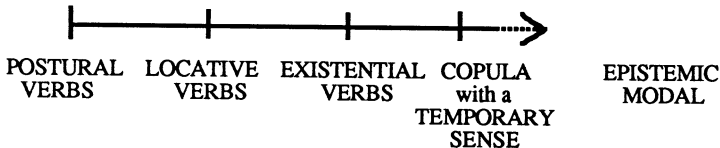
Figure 1: A Proposed Path of Semantic Evolution in Copulas

This view is supported using data from an informal cross-linguistic pilot study on copular construction. Evidence will be put forward to argue for the view that the 'temporary' meaning is an earlier stage of semantic development which leads into the epistemic or 'presuppositional' reading, and that this path is unidirectional. First, evidence is presented to show that postural and locative verbs are commonly involved in copular expressions. Next, the semantic notion of temporariness is discussed in relation to copulas, and the special affinity between locative expressions and temporariness is considered. Finally, arguments are offered for the hypothesis that the evolutionary path which is posited here is unidirectional.

The first claim that needs to be supported is that copulas do often find their source in verbs of posture or location. The pilot study, which looked at copular constructions in a convenience sample of languages, has revealed several other examples of copulas with connections to verbs of posture, in addition to the Spanish and Turkish cases already mentioned. Portuguese has a set of copulas that are identical to those found in Spanish. Irish and Scots Gaelic, Indo-European cousins of Spanish, also have a locative verb -- *ta* in Irish, *tha* in Scots Gaelic -- that is derived from the Proto-Indo-European root *sta-* , 'to stand'. This verb fits into a contrastive set with the copula *is* and yields 'temporary/ permanent' minimal pairs similar to those found in Spanish. The verb that is used for nominal predicates in Mangarayi is *ni* . The form is glossed 'to sit, be, exist', and it is noted that, although this form can be used as a general verb of location for most referents, it retains its original sense of 'to sit' to an extent that prohibits its use with nouns whose vertical extension is a more salient feature.

The languages just mentioned are cases in which a documented link to postural verbs was found in reference materials. The field of languages to be considered with regard to the locative- based historical sources for copulas is increased greatly when verbs of a more general locative meaning are included. In Sranan, for example, the copula *de* was reportedly a locative verb at an earlier stage of the language (cf. Favery, Johns and Wouk 1976). Example 3 shows that this copula is contrasted with another form, *na,* and that the distinction between the two is one that falls along the same lines as that expressed by Spanish *estar* and *ser*, with Sranan *de* expressing the temporary sense and *na* expressing an equation.

3. SRANAN
   a. Mi de botoman.   "I am a boatman (expresses the speaker's current occupation).
   b. Mi na botoman.   "I am a boatman (expresses the speaker's qualifications or professional capabilities.)

English provides another example. The suppletive past tense form of the English copula also derives from a locative Proto-Indo-European stem, *vas-* , meaning 'to dwell, to stay'. Another PIE root *men-* 'to remain' may be the source for the pair of copulas found in Balochi: *ynt,* used for singular definite subjects, and *ent,* used for plural subjects.

In addition to these examples, the copulas in several of the Chadic languages might also belong with this group. Frajzyngier 1986 presents a case for considering the copula *a* as being derived from a preposition in the proto-language. The form that is used in Mopun as an equative copula occurs in Bolewa as a locative copula, and in Fyer as both a locative and equative copula; thus, there is some strong motivation for considering it to have a locative source. However, given the persistent problems of classification that have been associated with the category of 'verbids' in African languages (cf. Lord 1973), where these elements display some behaviors that are verb-like and others that are like prepositions, an argument might be made for locative verbs as a historical source for them.

Language	Form of the Copula	Source
Spanish / Portuguese	estar	< L. stare 'to stand'
Irish / Scots Gaelic	ta / tha	< PIE *sta- 'to stand'
Mangarayi	ni 'to be'	< ni 'to sit'
Sranan	de	< general locative (< ? 'there')
English	was	< PIE *vas- 'to dwell, to stay'
Balochi	ynt ent	< ?PIE *men- 'to remain'
Chadic	a	< Proto-West Chadic a 'in, at'

Figure 2: Languages with Postural / Locative Sources for Copulas

Further evidence for copulas having a locative source might be taken from the synchronic situation that can be noted in many languages. It is not uncommon to find that there is a 'be'-like verb that is restricted to use with locative predicates. Such is the case in Mandarin, in which the particle *zai* can only be used with a locative complement. This form is distinct from the copula *shi*, which is used with nominal predicates. Korean, Lhasa Tibetan, and Igbo also have verbs which not only function as general locatives, but serve as existential verbs as well.

Mandarin	shǐ (copula)	Zhāng sān shǐ yi ge hùshǐ Zhangsan be one CL nurse 'Zhangsan is a nurse.'
	zài (locative particle)	Lǐsǐ zài hǎi-biān Lisi at ocean -side 'Lisi is at the oceanside
	yǒu (existential verb)	chéng li yǒu gōngyuán city -in exist park 'There is a park in the city.'
Korean	i-ta (copula)	Insu-nun haksaeng i-ta. Insu-TOP student be 'Insu is a student'
	iss-ta (locative/existential)	Insu-nun hakkyo-ey iss-ta. Insu-TOP school-in be 'Insu is at school.'

Lhasa Tibetan	reè (copula)	chạ ti tshāpo reè tea this hot be 'This tea is hot.'
	tuù (locative/existential)	thẹp tee tuù book here be 'The book is here.'
Igbo	bu (copula - identification)	ó bù ókú 3s be fire 'It is fire.'
	di (copula - description)	ó dĭ ókú 3s be fire 'It is fire-like; it's hot.'
	wa ...ni (locative/existential)	ó wà n'ìbè 3s be there 'S/He is there.' ó wà 3s be 'S/He exists; s/he is around.'
	na (locative, with human subjects; also glossed 'to sit')	ó nò nà Kánù 3s be at Kanu 'S/He is in Kano.'

Figure 3: Languages with distinct location copulas

With regard to those languages which use the same verb for location and existence, we note that Munro 1977 traces the copula in Pima, the verb-final particle -k , back to the existential verb *ka..[2] It is not a new observation that there is a certain degree of relatedness between the semantics of location and existence. Given the frequent identity in expressions for location and existence, it seems quite possible to view existence as a locative expression that is not specified for location, i.e. 'be located in the universe'. One might hypothesize that these cases of corresponding expressions for location and existence are, in some sense, a semantic bleaching, or perhaps more precisely a 'de-syntacticization', in that the existential expression develops out of a loss of the requirement for a locative complement[3].

An alternative scenario for development in these locational verbs might be that non-locative complements 'surrender to' the syntactic requirement for a locative complement that is imposed by the locational verb and mark the complement as a location. In the pilot study some languages were found that do follow such a strategy. Tamil requires that nominal and adjectival predicates in copular constructions be suffixed with the adverbial marker -aa when the copula iru- is used. Locative predicates do not require this marker, but use a locative marker in the same position. Hixkaryana similarly uses a 'denominalizer' postposition me with predicate noun complements of the copula. This postposition is also used as an adverbial clause marker. These facts seem parallel to the situation in the Celtic languages with the verb ta/tha , where a predicate noun phrase must be made locational by the use of a postposition.

Tamil	NON-LOCATIVE:
	A. atu nall - aa irukkutu
	   that good-ADVL be-pres-3s:neuter
	   'That is good.'
	B. avaru distrikt inspectar - aa iruntaaru
	   he district inspector-ADVL be-past-3s:hon
	   'He was a district inspector.'
	
	LOCATIVE
	raaman toottatti- le (irukkaraan)[4]
	Raman garden-LOC be-pres-3s:masc
	'Raman is in the garden.'

Hixkaryana	NON-LOCATIVE rowti me naha mosoni my brother DNML he-is this-one 'This is my brother.' LOCATIVE xamata ymo mkawo nehxakoni saraho rocky-island aug. on-top-of it -was manioc 'The manioc was on top of the rocky island.'

Figure 4: Languages with complements as adverbials/locatives

It does seem that there is compelling evidence for viewing locative and postural elements as potential sources for copulas. The question that arises once this point has been reached is whether there is legitimate evidence for suggesting that locative elements undergo a process of semantic bleaching that leads to their use as general copula-like forms. Some evidence toward an answer can be derived from the comparison of the use of copulas *estar* and *ser* in Spanish and Portuguese. Descriptions of both languages note the 'temporary' versus 'permanent' distiction for the use of the copulas. This semantic function appears to be most easily applied to adjectival predicates. The two languages differ, however, in their treatment of locative complements. In Spanish, only *estar* is regularly used with locative complements. In contrast, Portuguese allows both *ser* and *estar* to be used with locative complements, and furthermore, the 'temporary/permanent' distinction that is part of their use with adjectives is carried over into their use with locational predicates.

4. a. SPANISH
   i. Segovia <u>esta</u> en Espana   "Segovia is in Spain."
   ii. No <u>esta</u> en casa.   "He's not at home."

  b. PORTUGUESE
   i. A casa <u>e</u> no Flamengo   "The house is in Flamengo."
   ii. Joao <u>esta</u> em casa.   "Joao is in the house."

These facts provide some insight into the alternatives that might develop out of a single system. In both the Spanish and the Portuguese cases, the form *estar* has developed a sense of temporariness. This point is readily understood from the semantic distinctions that result from the choice of either *ser* or *estar* with an adjectival predicates. While adjectives as a class lend themselves to an interpretation in which the relative temporariness or permanance is important, just the opposite is true of nouns. If we consider nouns to be a semantic class that refers to stable and unchanging states or conditions, the inherent permanence of nouns is reflected by the fact that, in both Spanish

and Portuguese, only *ser* is used with predicate nouns. The difference between Spanish and Portuguese is the degree to which each has allowed the temporary sense of the copula *estar* to extend into the predication of location. Spanish *estar* seems to have retained a strong sense of location which requires that it be used with locative predicates. Portuguese *estar*, on the other hand, appears to have shifted its semantic sense so that its primary meaning component is more strongly a sense of temporariness (perhaps in contrast with the permanence implied by *ser*) which is carried over into locative predicates. Spanish may be moving to a similar stage. Butt and Benjamin 1989 relate that *estar* in Spanish is still the primary form for locative expressions; however, they also note that there is a tendency in the colloquial language to use *ser* with nouns that denote relatively more permanent features or fixtures.

There is some evidence that the locative copula in Korean *iss-ta* also has some connection to a meaning of 'temporary state'. The sentence in (5) below contains both the general locative verb and the copula *i-ta* . [5]

5. KOREAN
ku-nun   uysa - i̲ - ciman   cikum-un sensayng-ulo i̲s̲s̲-ta
3s-TOP doctor-COP-but   now -TOP teacher -as   be -IND

One possible reading for this sentence is, 'He is a doctor (by training), but right now he is (working as) a teacher.' Here again, the locative verb participates in the expression of a temporary state.

Another source for copular morphemes that has been discovered is deictic particles and personal pronouns (cf. Li and Thompson 1977). A trend that can be noted among the semantic distinctions that copulas express is that the 'temporary/permanent' semantic distinction is frequently noted in those cases where the copula is derived from a locative source, but copulas that are traced to deictic origins are more often noted to signal an identification relation rather notions of temporariness. Modern Hebrew is an example of a language that has copular forms that come from the third person pronouns. In Modern Hebrew, those clauses in which the copula is present are identificational; those that have no copula are predicational or descriptive (cf. Rapoport 1985). Perhaps a clearer illustration comes from Mandarin. According to Li and Thompson 1981, *you* predicates the existence of some noun at some location, while *shi* is used to characterize or identify the predicate noun. The distinction is demonstrated in the following examples.

6. MANDARIN
a. qiánmian yǒu   yi -ge huāyuán   'In front there is a garden.'
   in front exist one -CL garden
b. qiánmian shǐ   yi -ge huāyuán   'What's in front is a garden.'
   in front   be   one -CL garden
c. lǐmian yǒu shénme?   'What is inside?'
   inside exist what
d. lǐmian shǐ shénme?   'What is it that is inside?'
   inside   be   what

(6a) conveys information about the location of the garden; (6b) actually equates 'front' and 'garden'. This sentence would be used to indicate that either the only thing that is in front is the garden or that the addressee knows that something is in the front and merely needs to have it identified. This same distinction comes through in the possible answers to the questions (6c) and (6d). The question in (6c) could be answered with a phrase that means 'nothing'; however, such an answer would not be appropriate for question (6d). The explanation for this limitation is that *shi* presumes that some entity must be inside. The information that is required is what exactly that entity is.

The observation that these two sources tend to have particular usages associated with them has led to a hypothesis that those copulas found to convey a sense of temporariness are more likely to have their source with some locative element, and those that have a rigid identificational use are likely to have come from deictic sources. The basis for this hypothesis partially rests on the notion that the semantics of the source plays a deterministic role in the evolution of the grammatical item. In this case, the explanation is that location is not an inherent feature of a thing, i.e. an entity can be moved, re-located, and not have its identity changed at all. In contrast, deictic elements which can be used anaphorically to refer to some entity are functionally suited to the expression of identity. Their presence in an utterance is logically dependent on the fact that they have identical reference to some other entity in the utterance. [6]

Source	Characteristics	Examples
POSTURAL/ LOCATIVE VERB	- expression of temporary state - show a relatively full range of verb-like behavior - may be used as an auxiliary in complex or periphrastic verb constructions	Spanish (*estar*) English (*be*) Irish (*ta*) Korean (*iss-ta*) Igbo (*ino*)
DEICTIC PARTICLE or PRONOUN	- expression of identity - usually defective in the grammatical categories associated with verbs	Mandarin (*shi*) Modern Hebrew (*hu/hi*) Sranan (*na* < *da* "*this*")

Figure 5: Two sources for copulas and their general characteristics

In the course of doing the pilot study, two basic patterns were noted with respect to the expression of temporariness in copulas. One typical system is exemplified by the system in Spanish. There are two verbs, one of which denotes a temporary condition while the other denotes a permanent state. This kind of dual system also occurs in Irish, Scots Gaelic, and Sranan. Old English also maintained a similar distinction using two separate verbs 'to be'. *Beo-* was used to express habitual occurences, while *wes-* expressed identity or eternal truth.

Traugott 1972 links this last historical fact to the invariant *be* of Black street speech. Although the use of invariant *be* is not fully understood, a number of researchers (Fasold 1969; Rickford 1974, Baugh 1983) have noted that it frequently appears with adverbial indicators of distributive habitual activity, but less frequently with non-habitual phrases, and never in cases of identity. In these latter two aspects a zero-form copula is generally used. This exemplifies the second pattern that can be found for the marking of temporariness. In these cases there is a single copula, and its optional appearance is a signal of the relative temporal limitations of the predication. Another example of a language that employs this strategy is Kobon. In present perfect tense the copula is generally absent unless the speaker regards the feature as being of a temporary nature. In that case, the copula may be overt.

7. KOBON
    a. Nipe   Kaunsol   (m+d - öp)   'He is the councillor.'
       3s   councillor   be - perf3s
    b. Nipe b+ majo (*m+d - öp)   'He is a mature man.'
       3s   man   mature be - perf3s

It is also worth noting that the Kobon copula *m+d* is identical to the marker of habitual aspect. Another example of a language that makes a temporal distinction based on the use

or non-use of the copula is Turkish. The copula *-dir* is used in statements that have general validity. The same statement without the copula refers to the current moment.

8. TURKISH
   a. Atlar   tembel-dir   'Horses are lazy (a general statement of fact).'
   b. Atlar   tembel   'The horses are lazy (at the moment)'

The two methods of 'temporary/permanent' distinction are summarized in Figure 6 below.

Language	Form	Example	
SPANISH	estar	Juan esta enfermo 'Juan is sick.'	
	ser	Juan es enfermo 'Juan is sickly.'	
IRISH	ta	ta an páipéar ban 'The paper is white.' (i.e. blank; not written on)	**2**
	is	is bán an páipéar e 'The paper is white.' (i.e. its whiteness is inherent)	
SRANAN	de	Mi de botoman 'I am a boatman.' (refers to the speaker's current occupation)	**F**
	na	Mi na botoman 'I am a boatman.' (by training)	**O** **R**
OLD ENGLISH	beo-	...þonne þaer bið man dead... '...whenever there is a dead man'	**M** **S**
	wes-	..of Danai þaere ie, seo is irnende of norþdaele... 'from Danai that river which is running (=which runs) from northern part'	
BLACK ENGLISH	be	Sharon be neat eriday	
	Ø	They in the club. (at that time)	**1**
KOBON	m+d	Nipe Kaunsol m+d-öp 'He is the councillor.'(for the time being)	**F**
	Ø	Nipe Kaunsol 'He is the councillor.' (no time restriction implied)	**O** **R**
TURKISH	-dir	Atlar tembeldir 'Horses are lazy.' (generally)	**M**
	Ø	Atlar tembel 'The horses are lazy.' (right now)	

Figure 6: Expression of 'temporary'/'permanent' distinction

To summarize the observations that have been made so far, there is good evidence that the source for the copula morpheme, in at least some languages, can be found in locative or postural verbs. I've provided evidence from historical reconstructions. I've also tried to support this claim by providing evidence from synchronic grammars that shows some degree of connection between copulas and expressions of location. Finally, I've attempted to show that cross-linguistic observations demonstrate a particular affinity between the expression of temporariness and the notion of location, and to provide some reason why that affinity may exist.

Let me now return to the observations that began this presentation. In the introduction to this presentation , I suggested that it is possible to relate the temporary sense that is expressed in Spanish *estar* and the presuppositional sense that is expressed in

Turkish -*dir*. The notion of a 'path' that I am invoking here is a particular instance of the general process of 'grammaticization' that is outlined in Bybee 1985. The term 'grammaticization' refers to the development of grammatical morphemes from full lexical items through parallel and gradual processes of phonological erosion and semantic generalization. The process is constrained by mediating principles that Bybee labels 'relevance' and 'generality'. Of particular interest for the considerations at hand is the principle of generality, which maintains that the sources for grammatical morphemes must be sufficiently general in their semantic content to allow them to be combined appropriately and legitimately with a wide range of elements. This factor, however, must be counter-balanced with a general necessity that describes the communicative intent that characterizes any utterance. That is, any lexical item that is to serve as a source must have enough semantic content to make it useful in an utterance. It must contribute something to the meaning of the utterance in order to assure that it is used frequently, which in turn will lead to further phonological and semantic reduction, all of which ultimately leads to its development as a gram. The effect of this principle of generality is to limit the stock of potential sources for grammatical markers.

Grammaticization research also suggests that the semantics of the lexical source plays a deterministic role in the nature of the gram that ultimately develops out of the source. Bybee and Pagliuca 1987, for instance, gives evidence for six specific semantic sources for future tense morphemes. Heine and Reh 1984 presents material that also demonstrates the regular occurence of some specific lexical sources and associated endpoints for grammatical markers in African languages. The strongest stance on this issue would be to suggest that the original semantics of the lexical source may not be lost entirely in the process of bleaching that accompanies grammaticization. The semantics of the source is recoverable to a certain extent, and that recoverability comes through in restrictions and apparent anomolies in the synchronic behavior seen in some grams.[7]

The principle of generality determines that only a small number of lexical items are potentially available to undergo grammaticization. The idea of a kind of semantic determinism operating in the process of grammaticization suggests that the members of this limited set are put on a particular path by their original semantics. Taken together, these principles provide a more compelling background for considering the etymological parallel that exists between the Spanish and Turkish copulas. The fact that these two morphemes share a source, as well as a grammatical function, provides strong motivation for relating them both to a single evolutionary path of semantic development. The final question that needs to be addressed is how the meaning of 'temporary state' and 'presupposition' might be related. I suggest that it is possible to see presupposition as an extension in the scope of 'temporary state'. That is, the notion of 'temporariness' is not simply applied to the action of the predicate, but is extended over the entire assertion. In the final stage of its evolution, the Turkish copula takes the entire clause within its scope, and rather than asserting that the state which is predicated of the subject holds temporarily, it presents the nature of the assertion itself as temporary. The proposed stages of development are schematized in Figure 7.

STAGE I.   The children stand in the garden.

STAGE II.   The children $\begin{bmatrix} \text{are in the garden.} \\ \text{GENERAL} \\ \text{LOCATION} \end{bmatrix}$

STAGE III.   The children $\begin{bmatrix} \text{are in the garden.} \\ \text{TEMPORARY} \\ \text{LOCATION} \end{bmatrix}$

STAGE IV.
$$\begin{bmatrix} \text{The children are in the garden.} \\ \text{TEMPORARY} \\ \text{ASSERTION} \end{bmatrix}$$

(...i.e.,'I temporarily assert, until presented with evidence to the contrary, that the children are in the garden.')

Figure 7: Proposed stages on the path of semantic development in the Turkish copula

An iconic measure for this view of the situation is provided by the example in (9).

9. TURKISH
   a. bahçede   'in the garden'
   b. bahçedeler   'They are in the garden.'
   c. bahçededirler   'They are in the garden (EMPHATIC) *or* 'They are surely in the garden.' (PRESUPPOSITION)
   d. bahçedelerdir   'They are surely in the garden.' (PRESUPPOSITION)

Note that (9c) and (9d) contain all of the same morphemic elements, differing only in the ordering of the plural marker *-ler* and the copula *-dir* . These two elements can shift positions; however, the utterance is unambiguous only when the enclitic *-dir* is in the final position. It seems that only when filling this final slot, taking the rest of the proposition within its domain, can the copula clearly function in its epistemic sense.

Another indication of the stage to which the Turkish copula has evolved may be taken from the fact that it is now freely affixed to verbal forms to indicate the speaker's presumption of the truth of the statement. In other words, it is not simply used to equate or identify a third-person subject with a complement, nor to add emphasis or a sense of supposition to such statements, but also to add an epistemic sense to any utterance. The examples given in (10) below show that the copula is no longer restricted to use with third person subjects. It can now be used with all forms of the verb, no matter what the person/number marking is, and its function in this use is to add an epistemic modal flavor to the meaning of the utterance.

10. a. şair-im   'I am a poet.'
    b. şair-im-dir   'I am surely a poet; I think that I must be a poet.'
    c. biliyor-sunuz   'You all know .'
    d. biliyor-sunuz-dur   'You all surely know ; I presume that you all know.'

The final point which requires some empirical support is the claim that the development from 'temporary state' to 'presupposition' is unidirectional. As stated earlier, a key component in the process of grammaticization is semantic generalization. The productive use of *-dir* as a modal element with all predicates, regardless of their person/number marking, suggests that it has undergone a semantic bleaching. That component of meaning that at one time required that *-dir* be used only with third person subjects has been lost.

Another bit of evidence that favors the unidirectionality hypothesis is the fact that, among the languages that were looked into in the pilot study, there were only three languages in which the copula was a form that was bound to the stem: Hebrew, Buriat and Turkish. The first two cases appear to have sources in pronouns, and their affixal form might be seen as a natural development in the direction of person agreement markers. In contrast, Turkish was the only language with a verbal source for a copula that is bound. Bybee's view of grammaticization posits a correlation between semantic generalization and phonetic erosion, with more generalized meaning signaled by greater phonetic reduction

and fusion. The unusual nature of the Turkish form suggests that it has travelled further along the evolutionary path than the non-bound copular morphemes.

Finally, I would like to propose that the notion of temporariness is logically precedent to the notion of presupposition. I am certainly not the first to suggest that the directionality that has been observed in language change is driven by a transfer from the concrete experience to the abstract (cf. Traugott and Konig 1988 for a summary and discussion). The understanding of temporariness, of states reaching an endpoint, would seem to be a concept that is readily accessible to humans on a concrete level. On the other hand, the epistemic notion of possibility seems more abstract in that it requires the speaker to make use of the concrete artifacts of a situation to infer some outcome.

One problem persists in this analysis of the Turkish epistemic-sense copula. That is the fact that the 'temporary state' meaning which I have posited as an intermediate stage of development in the general schema does not appear to be a component in the meaning of the Turkish copula. Indeed, there is a counterintuitive clash of meanings when one considers the fact that, in Turkish, the copula contrasts with a zero and gives the sense of 'general validity' (see the examples in Figure 6 above), and at the same time has the epistemic use that I have attributed to a 'locative-to-temporary' path of semantic evolution. At present I can offer no firm resolution of this puzzle, except to revise my original conception of the path to view the development of the 'presupposition' sense. Under this revision, the epistemic sense is not viewed as a direct development of the copula, but rather a development specifically tied to the 'temporary' semantics. There is a 'fork in the road', as it were, and the single form can continue to develop on the path towards a general copular use, as well as functioning as a modal, a sense that is retained from an earlier stage of its evolution.

POSTURAL VERB   LOCATIVE VERB   COPULA with a TEMPORARY SENSE >>>>>>>>>>>>> PRESUPPOSITION   GENERAL COPULA
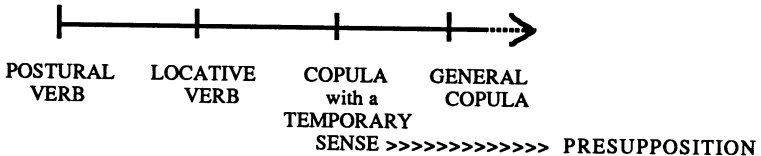
Figure 8: A Revised Path of Semantic Evolution for Turkish *-dir*

Footnotes

[1] I would like to thank Joan Bybee and William Pagliuca for their guidance and comments on this work. I would also like to thank Soon Ae Chun for helping me to clarify some of my ideas, and to express appreciation to Lloyd Anderson, Susan Herring, Gary Holland, Leslie Saxon and Eve Sweetser for their comments. I, of course, am responsible for any errors contained herein.

[2] In Munro's reconstruction, the marker *-k* does not act as the copula by itself. Rather, it is supported in this function by the impersonal pronoun **pul* .

[3] Lyons 1968 appears to suggest just the opposite direction of development; i.e., location is derived from existence, in the sense that 'anything that exists must exist in some place.'

[4] Susan Herring has informed me that the optional presence of the copula in this sentence renders a change in its meaning. The sentence with the copula implies simple location, whereas the sentence without the copula implies a more permanent sense, i.e. 'Raman is planted in the garden.'

5 I am indebted to Soon Ae Chun for providing this example.

6 There is a point of logical convergence for these two points. That is that identificational copulas might also lend themselves to an interpretation of permanance of timeless truth. In this regard, Turkish might be a case that bears closer scrutiny since the forms of the copula for first and second person are former personal pronouns.

7 I believe that such a position is suggested in Claudi and Heine (1985) in which insight into some anomolies in the grammatical treatment of inalienable possessions in Ewe are explained by appealing to the meaning of the source item for the genitive item. William Pagliuca (p.c.) has articulated similar views, suggesting a 'windowing effect' through which past semantics are synchronically accessible.

REFERENCES

Language References
ASHER, R. E. 1985. Tamil. London: Croom Helm.
BARKER, Muhammad Abd-al-Rahman and Aqil Khan Mengal. 1969. A Course in Baluchi. Montreal: Institute of Islamic Studies, McGill University.
BAUGH, John. 1983. Black Street Speech: Its History, Structure and Survival. Austin: University of Texas Press.
BUTT, John and Carmen Benjamin.1989. A New Reference Grammar of Modern Spanish. London: Edward Arnold.
DAVIES, John. 1981. Kobon. Lingua Descriptive Studies, volume 3. Amsterdam: North-Holland Publishing Company.
DERBYSHIRE, Desmond C. 1979. Hixkaryana. Lingua Descriptive Studies, volume 1. Amsterdam: North-Holland Publishing Company.
GOLDSTEIN, Melvyn C. and Nawang Nornang. 1970. Modern Spoken Tibetan: Lhasa Dialect. Seattle: University of Washington Press.
LEWIS, G. L. 1967. Turkish Grammar. London: Oxford University Press.
LI, Charles N. and Sandra A. Thompson. 1981. Mandarin Chinese. Berkeley and Los Angeles: University of California Press.
MERLAN, Francesca. 1982. Mangarayi. Lingua Descriptive Studies, volume 4. Amsterdam: North-Holland Publishing Company.
SWIFT, Lloyd B. 1963. A Reference Grammar of Modern Turkish. The Hague: Mouton.
TRAUGOTT, Elizabeth C. 1972. A History of English Syntax. New York: Holt, Rinehart and Winston.
UNDERHILL, Robert. 1976. Turkish Grammar. Cambridge, Massachusetts: MIT Press.
WELMERS, William E. 1973. African Language Structures. Berkeley and Los Angeles: University of California Press.

General References

BYBEE, Joan L. 1985. Morphology: a study of the relation between form and meaning. Amsterdam: John Benjamins.
BYBEE, Joan L. and William Pagliuca.1985. Cross-linguistic comparison and the development of future meaning. in Historical Semantics and Historical Word Formation, ed. by J. Fisiak, pp. 59-83. Berlin: de Gruyter.
BYBEE, Joan L. and William Pagliuca.1987. The evolution of future meaning. in Papers from the VIIth International Conference on Historical Linguistics, ed. by A. G. Ramat, O. Carruba and G. Bernini, pp. 109-22. Amsterdam: John Benjamins.

CLAUDI, Ulrike and Bernd Heine. 1985. From Metaphor to Grammar: Some Examples from Ewe. Afrikanistishe Arbeitspapiere #1, pp. 17-54.

COMRIE, Bernard. 1976. Aspect. Cambridge: Cambridge University Press.

FASOLD, Ralph W. 1969. Tense and the form *be* in Black English. Language 45, pp. 763-76.

FAVERY, Margot, Brenda Johns and Fay Wouk. 1976. The Historical Development of Locative and Existential Copula Constructions in Afro-English Creole Languages. in Papers from the Parasession on Diachronic Syntax, ed. by S. Steever, C. A. Walker and S. S. Mufwene, pp. 88-95. Chicago:Chicago Linguistic Society.

FRAJZYNGIER, Zygmunt. 1986. From Preposition to Copula. in Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society. Berkeley: Berkeley Linguistics Society.

HEINE, Bernd and Mechthild Reh. 1984. Grammaticalization and Reanalysis in African Languages. Hamburg: Buske.

LI, Charles N. and Sandra A. Thompson. 1977. A Mechanism for the Development of Copula Morphemes. in Mechanisms of Syntactic Change, ed. by Charles N. Li, pp. 419-44. Austin: University of Texas Press.

LORD, Carol . 1973. Serial Verbs in Transition. Studies in African Linguistics 4. pp. 269-96.

LYONS, John. 1968. Introduction to Theoretical Linguistics. Cambridge: Cambridge University Press.

MUNRO, Pamela. 1977. From Existential to Copula: The History of the Yuman BE. in Mechanisms of Syntactic Change, ed. by Charles N. Li, pp. 445-90. Austin: University of Texas Press.

RAPOPORT, Tova R. 1985. Copular Constructions in Hebrew. in CLS 21: General Session, ed. by W. Eilfort, P. Kroeber and K. Peterson, pp. 354-70. Chicago: Chicago Linguistics Society.

RICKFORD, John R. 1974. The insights of the mesolect. in Pidgins and Creoles: Current Trends and Prospects, ed. by David DeCamp and Ian F. Hancock, pp. 92-117. Washington, D.C.: Georgetown University Press.

TRAUGOTT, Elizabeth C. and Ekkehard Konig. 1988. The Semantics-Pragmatics of Gramaticalization Revisited. ms of a paper prepared for the Symposium on Grammaticalization; Eugene, Oregon. 1988.

Therapeutic Flouting: Strategic Uses of Metaphor in Psychotherapy

Kathleen Ferrara
Texas A&M University

Rather than focusing on the Legacy of Grice, this paper addresses the challenge to Grice. I see this challenge in two areas. First, the Cooperative Principle and its maxims are formulated to deal with conversation. But far more common than conversation is Language at Work, the situated use of language in daily work, including job interviewing, opinion polling, the making of travel plans, and talking about problems in psychotherapy. We must ask, 'Do the maxims which comprise the Cooperative Principle apply equally well to these everyday forms of talk? Second, we must find out if the maxims hold up in cases of extended, naturally occurring language as well as they do in our prized isolated sentences. To do this, we need actual samples of real interaction, to add to decontextualized or fabricated examples already in the literature.

The focus of this paper is on the actual and situated use of metaphor in Therapeutic Discourse. Thanks to Labov and Fanshel (1977) we know many important things about this speech event, one which has only emerged in this century. One speech act in this type of event that has not received a great deal of attention is the use of figurative language. A puzzle is why, in a speech event where truth telling is a norm, do we have flouting of Grice's Maxim of Quality (Do not say that which you believe to be false.)? Even the Maxim of Relevance appears to be violated in metaphor. The answer lies in the observation that serious messages are often communicated in speech play. As Sherzer (1982) points out, communicating obliquely is less threatening. It often suits people, especially troubled people, to take advantage of what Martinich (1984) refers to as 'metaphor indeterminacy,' graphically represented in (1)

When one speaker declares to another that A is B or A is like B the extent of the overlap (or similitude) is not declared. The first speaker merely asserts that two things are related; the second is charged with figuring out how two things are related. Martinich (1984:284) states that this indeterminateness is what makes metaphor salient, because it requires effort to resolve. The vibrancy of a metaphor's unresolved tension can be seen in (1) where the reverberations are indicated with dotted lines.

(1)
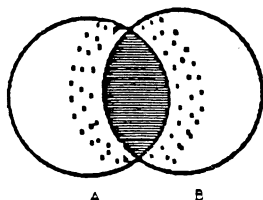
Metaphor Indeterminacy

In a metaphor proposed by Jake, one of the psychotherapy clients in the current corpus, 'marriage is a cocoon,' for example, it is up to the recipient to decide if the interlocutor means that it is a) a safe place, b) a place to develop and grow, c) a place of confinement, d) a place to escape from, or e) all of these or others.

I claim that the interpretation of metaphor can never be exhaustively arrived at by semantic analysis because of two considerations. First, the reverberations are not finite but open-ended. Even long after utterance of a metaphor, new reverberations can be felt, new connections can be seen. Second, interpretation is always contextually based--dependent on the situation, the speakers, and the previous discourse which give rise to the figurative expression. For these reasons, a focus on actualized metaphor, metaphor in use, is more revealing than a static semantic analysis. This calls for a discourse approach to metaphor.

Despite Sapir and Crocker's (1977) call for further examination of the social uses of metaphor, we have only a handful of such studies. Crocker (1977) notes the creativity involved in applying even a conventional metaphor at the right moment:

> Even if a metaphor is commonly known it still must be applied at a particular social moment. This involves a sense of aptness, of timing and gestalt perception of context which is just as creative as coining a new phrase.

There is a vast literature on metaphor from differing perspectives. Psycholinguists such as Ortony and Osgood; anthropologists (e.g., Fernandez, Howe, Leach and Sherzer); philosophers and linguists (e.g. Verbrugge, Lakoff and Johnson) have provided insights on the nature of metaphor. The functional, interactional approach to the study of 'what social ends...they serve' (Sapir and Crocker 1977:33) and how the intentions of non-literal speech should be understood is the closest in orientation to that of the present study.

Previous interest by psychologists in metaphor has principally been in the realm of comprehension or processing. Only a handful of research targets the therapeutic uses of metaphor. Lenrow's (1966:145) claim that discussion of the role of metaphors in individual therapy is rare is still true nearly a quarter of a century later, despite the fact that Pollio and Barlow (1975:112) argue that most experienced therapists know that figurative language plays a significant role in the therapeutic process. For example, Pollio and Barlow (1975) found an average of three metaphors per 100 words in a single hour of therapy. Pollio, Barlow, Fine and Pollio (1977) hypothesize on the basis of four independent coders' judgments that 'insight' occurs simultaneously with or adjacent to novel metaphorical utterances. Lenrow (1966:147) also sees metaphor as valuable in developing insight. He states

> The half playful style of metaphorical descriptions permits the psychologist to communicate about important aspects of the client's behavior while ostensibly talking about a class of events very different....In psychotherapy, this is sometimes termed communicating with the client in an 'area of displacement.'

For this study I will draw examples from 48 hours of tape-recorded individual psychotherapy sessions between 6 different therapists and 10 individual clients. Four of the therapists in the study are experienced clinical psychologists, one is an experienced psychiatric social worker, and two are clinical psychologists in training. The clients are both male and female, middle and lower class, with presenting problems ranging from alcoholism, depression, low self-esteem, physical and sexual abuse. The data was collected with advance written permission in Texas and Oklahoma. The method I used was to tape-record six consecutive sessions, beginning, in most cases, with the first session of therapy so that I could have access to preceding and succeeding discourse and all associations. Only the client and therapist were present at the time of recording.

The purpose of the paper is to explore four basic ways in which metaphors can be received in therapeutic discourse and to illustrate with actual short examples the manner in which therapists receive client-initiated metaphors. These four ways are shown in (2).

(2) Four Ways in Which Metaphors Can be Received in Psychotherapy
    I. Comprehension without Comment
    II. Misunderstanding, Failure to Comprehend or Get the Point
    III. Ratification by Comment, Repetition, or Use of Associated
          Word in Later Discourse
    IV. Extending the Metaphor Jointly, Corroborating by Helping to
          Construct a Chain of Collocational Cohesion

First, therapists can comprehend a client-proffered metaphor but let it pass uncommented upon, without further attention. Second, therapists can fail to comprehend a metaphor and miss the point. Such misunderstandings can be costly. If the misunderstanding itself becomes noticeable, attempts to rectify it can ensue. (As Lentine (1988) shows, attempts at clarification can consume otherwise valuable time.) An example of a misunderstanding is given below. Third, mental health specialists can show appreciation or otherwise 'reward' the client by ratifying the metaphor. Ratification can take several forms, ranging from a comment (e.g. 'That's right') to repetition of the client metaphor, to use of a semantically entailed or associated word or phrase in later discourse. Therapists can also take up another similarity or dimension of a metaphor, one not brought out by the client. This is illustrated below. A fourth way in which therapists can corroborate the aptness of a metaphor is to help construct a chain, to assist in weaving an elaborate web of correspondences which tease out additional ramifications and add new dimensions. Rather than presenting metaphors of their choosing (cf. Gordon 1978), therapists can try to emphasize the raw material presented by clients, and, if possible, use, the lead established by them to spin out further connections. In this fourth manner, they can exploit a natural aspect of langauge, lexico-semantic cohesion, as a strategy to densely layer semantic associations in jointly constructed extended metaphor. They can use what Halliday and Hasan (1976) designate as **collocational cohesion**, the cohesion that results from the co-occurrence of lexical items that are in some way or other typically associated with one another, such as synonyms, antonyms, superordinates, synechdoche (part-whole), hyponyms or co-hyponyms of the same superordinate term (i.e., both members of the same more general class, such as *chair...table*--both hyponyms of *furniture*). An example of an extended case will follow.

To illustrate various ways in which metaphors can be received, a sequential approach to metaphor in discourse will be used. However, it is also informative to study metaphor from a distributional approach. In the current corpus, both metaphors *for* and *in* therapy are revealing. For example, the client Norma, when interviewed for this study and asked what therapy was like, resorted to metaphor, stating, *'Therapy is a dance. First my therapist takes a step, then I take a step and then we're moving together. It's beautiful.'* To express the unfamiliar she uses the familiar.

Lakoff and Johnson (1980:47) comment that a pervasive structural metaphor is *'ideas are food.'* All of the therapists in the corpus describe therapy as a process. One suggests, by means of numerous cooking metaphors, that the process is similar to the creative process of food preparation.

(3) (session initial)
-*What you got cooking* today?

(speaking of a lengthy string of dreams)
-That does suggest that *something's on the front burner.*

-Well, I'm struck, not so much with the details of the themes, although I hear the sexual violence in that, but uh, but with the fact that they're coming and they're *perking up* lots of rich detail, lots of complications...

-Sometimes so much *perks up, bubbles to the surface.*

Metaphor establishes a special relationship to truth: it approximates truth. Because the dimensions established only touch upon but do not coincide with the truth, often several metaphors for a concept are necessary. Multiple metaphors jointly contribute to forming a coherent picture. A tripartite formulation which attempts to capture the dimensions of an abstract concept like anger is indicative of what Lakoff and Johnson (1980) see as structural metaphors. For example, one very angry client in the corpus who was abused as a child and feels 'in danger of exploding' formulates three different metaphors for anger:

(a) *Anger is Poison*   (b) *Anger is Fire*   (c) *Anger is a Weapon*

(4) (Re (a) *Anger is Poison*)

Sharon: It's just a lot of *pus* and *poison* getting over everything.

Marian: You let your guard down and what came *squirting* out?
Sharon: (2) Anger ((laugh))
Marian: Yeah.

(Re (b) Anger is Fire)

Sharon: Yesterday afternoon *reignited* it. I don't like how much *fire* there is in me.
Marian: See that's what I think you're afraid of finding in you is all . [that *fire*] all the anger and *the flames*...
Sharon: [that *fire*]

Marian: No one would ever dream that beneath that calm, cool exterior (1) there raged *a fire*.

(Re (c) Anger is a Weapon)
(explaining why she doesn't often express anger to her mother)

Sharon: ((laugh)) You don't like to *bludgeon* people either. I don't do that very often.

Marian: It could be that she see::s your anger . or senses it and keeps you know a ((laugh)) bo:dy between the two of you so . to deflect *the missles* . or it could be that she needs the distance too . to protect herself.

Often multiple metaphors are necessary to capture the magnitude of a concept. The overlapping edges help define it.

There is a remarkable consistency both within and across clients in the particular domains they favor for figurative expressions. For example, the client Wilma consistently, through six sessions, speaks of aridness, suffocation, drying up, and the stifling aspects of the hometown she has fled. Likewise, two-thirds of the clients refer to erecting *walls* in discussing barriers to communication.

With this distributional account of metaphor, we turn now to a sequential analysis of actual short samples of interaction involving metaphor and the various types of reception they have. The first sample will be a microanalysis of a Misunderstanding. Despite the fact that therapy is considered an appropriate place for expressive language, not every excursion into metaphor is successful. Consider the apparent misunderstanding shown in (5) in the segment from Jake and Bonnie which arises from confusion over Jake's metaphoric use of *hill* in the statement, 'Well, it's not going to be much longer and I'm going to be *over the hill* ((mumble)) and it's gonna be *downhill*.' As background, Jake is a 36 year old computer salesman with an engineering degree whose abusive wife has recently filed for divorce. Bonnie is a 26 year old clinical psychologist in training. The discrepancy in their ages may be a key to their interaction.

(5) **Metaphor Misunderstanding**

Jake: It's always when I've sold and I'm having you know a bad month in sales or maybe a bad two or three months on sales.
Bonnie: Mmhmm
Jake: Uh the only way I could ever get out of that was just to work twice as hard at it.

Bonnie:	( )	
Jake:	I think I'm trying to use the same thing again.	
Bonnie:	So you're having bad times so you're going to work extra hard and maybe make up for it.	
Jake:	Mmhmm. ((laugh)) It's not looking good! I've had a bad first quarter as far as sales	
Bonnie:	Yeah	
Jake:	and it's going into the second quarter. Uh, you know, it's all looking real bad and I'm saying, 'Well, it's not going to be much longer and I'm going to be *over the hill* (((mumble)) °and it's gonna be *downhill*.° I don't know. I'm just trying to hang on till that time.	
--> Bonnie:	When you say "*over the hill*," what do you, what do you mean?	
Jake:	Oh, probably feeling sorry for myself. You know, just thinking, everything is happening, everything is going wrong with me.	
Bonnie:	Mm	
Jake:	Well, I think there's good things if you can-	
--> Bonnie:	So *the hill* really is not ba:d ( ) [but good.]	
Jake:	[Yeah.] About the time you think you're at *the peak of it*, well, some drunk Indian backed into my car door and caved it in.	
Bonnie:	Oh.	
Jake:	So uh I don't know if I told you . and I didn't want to turn him in, you know and call the police at that time, didn't wanna give any trouble...	

The therapist appears to realize there is a mismatch in interpretations when she says, at the first arrow, 'When you say "over the hill" what do you, what do you mean?' The client responds as if she had said, 'Why do you say that?' rather than 'In what sense do you mean that?' The indication that her interpretation of *over the hill* is pejorative and does not match that of the client is in the question, at the second arrow, 'So *the hill* is really not ba:d ( ) but good?'

Baed on his connections with the italicized portions, *peak* and *downhill*, the client appears to use *hill* to signify life's struggles, as in the myth of Sisyphus, seized by Camus as a metaphor for existentialism, because he makes reference to how he has worked hard, expected success, but has encountered repeated setbacks at work (and in the marriage, in previous discourse). This example of miscommunication is caused by faulty uptake of the *hill* metaphor. There are two different interpretations. One is the client's use of the metaphor; the other is the therapist's idiomatic reading. Jake's understanding is that 'Life is a hill. Uphill is hard, but downhill is easy. You need to try to get to the peak. Like Sisyphus, you may experience setbacks, but you keep trying.' On the other hand, Bonnie's interpretation is centered on the idiom 'over the hill,' meaning 'past one's prime, near 40, on the wane.' Ortony (1980) argues that processing of idioms is more rapid than that for metaphor. In terms of Burbules, Schraw and Trathen's (1989) discussion of the difference betwen idiom and metaphor, the therapist here has approached 'over the hill' as requiring the retrieval of institutionalized meaning (idiom), whereas the client has generated meaning, inviting figuration (metaphor). Because 'figuration occurs interactively' (Burbules et al. 1989:107), and the figuration is one-sided here, the metaphor is unsuccessful. Crucially, in this example, the two interlocutors never do resolve their misunderstanding.

Far more typical of the successful therapeutic encounters in the corpus, however, is the frequent occurrence in which therapists, moments or minutes after a client's initial

metaphor, bring that metaphor back into focus, either by repetition, synonym, or with semantically associated terms. This form is called Ratification. What is communicated by such ratifications is that the client's word choice has been attended to and is deemed potent. When a client's own figurative language is held in consciousness for a while and then reemployed by a therapist, the message is that metaphor is a valuabe resource for understanding or insight.

An extract which illustrates this pattern, shown in (6), is taken from the second session between the therapist Ralph and Lana, a recovering alcoholic who is separated from her husband and has two teenagers at home. She has been sober for one year but is subject to moods where she feels overwhelmed, out of control. She describes below how, after one bout, she had a blow-up with her children but is now developing a new coping strategy for mood changes.

(6) Ratification of Metaphor

RL(2)32

Lana: And the next time *I felt it coming I could feel it*
            coming and I just kinda (3) did anything I could
            to get the pressure off. You know, I called people, I
            got my body out of the house and you know I figure,
            I hope, *if I can see it coming,* I can look at it and just
            go, *"There is an insanity tidal wave coming.* I'm going
            to take this and myself out of your [the kids'] face so
            you won't get it, because you don't deserve it but *it's*
            coming." ((laugh))

Ralph: Mmhmm

Lana: I want to do that.

Ralph: Yeah, anything that you can do

Lana: Mmm

Ralph: that's active .
            to cope with that. That means calling your friends,
            turning to an enjoyable, uh, diversionary activities,
            going to a movie.

[In intervening two minutes Lana continues, describing another such 'attack' after the build-up of pressure in which she felt out of control.]

Ralph: ...   but you're getting overwhelmed with it

Lana: Uh-huh

Ralph: to be perfect. And you've finally caught yourself and
            realized, 'Wait a minute. This is crazy. This is
            insanity.' You called first one-

Lana: And it's gonna make everything crazier.

Ralph: Right. *It, the tidal wave, it just grows bigger and*
            bigger and bigger. It keeps growing. Uh, this ( )part,
            the thing about the guardian angel

Lana: Oh

Ralph: Can you
            tell me a little bit more about that?

In the segment above Ralph ratifies Lana's metaphor for insanity, *the tidal wave,* by referring to *the tidal wave* two and a half minutes later. He also uses one other ratification device. The indeterminacy feature of metaphor discussed previously, its virtually inexhaustible reverberations, comes into play in this example. Notice that Lana emphasizes five times (in italics) that *the tidal wave 'is coming.'* Lana figuratively conveys the natural, overpowering force she experiences at times. Her phrasing, *'feel it*

coming' makes prominent both the inexorable and the unpredictable qualities of her emotions. She can't stop them but she can get out of the way. However, the aspect which the therapist Ralph emphasizes is different. He points out that *the wave* gathers size and momentum as it approaches. *'It just grows bigger and bigger and bigger,'* he says. These slightly different aspects are nonetheless complementary, and, of course, there are other aspects waiting to be taken up in subsequent discourse. Aftell and Lakoff (1985) claim that by such acceptance and extension of client's original uses of language, therapists can build rapport and facilitate further client self-expression.

Nonetheless, even greater degrees of ratification are possible. The final example, shown in (7) and schematized in (8) and (9), illustrates how client and therapist can jointly enhance meaning and collaboratively construct an extended metaphor with unusual power and relevance for the client. I claim that collocational cohesion (Halliday and Hasan 1976), an inherent characteristic of language, is purposely stretched out and exploited in this discourse segment. In the example below the play aspects of lexico-semantic repetition are very noticeable. The jewel-like nature of this extract illustrates what Chafe, Tannen, and Sherzer see as the verbal art and the poetics of everyday discourse.

As background to the segment in (7), Howard is a client in his mid-thirties who is talking with Judy, his therapist, in their third session. It is relevant to note that Howard has a college degree, has served in Viet Nam, but for a while has been underemployed, working as an orderly in a hospital. Crucial to this segment, a month prior to the session he was fired from the hospital because he was suspected of taking drugs which were missing. He maintained that he did not take the drugs and has since been reinstated in his job, with apologies. He has experimented with several drugs. He considers the period in which he fought for reinstatement as a challenging period. In this segment a structural metaphor is proposed by the client.

Floating	*down the River*	*in a Great Ole Big Barge.*

represents

Drifting	through Life	in a Low Level Job

Both client and therapist weave an elaborate web of extensions. Notice the cohesive items shown in italics.

(7) Joint Construction of Extended Metaphor
JH(3)1
1. Judy: When you have a problem, what do you do with it?
2. Howard: I usually let it be a problem. I don't usually do any-
3. thing much or I . I was thinking about that the other
4. day.
5. Judy: Does the problem go away if you don't do anything
6. about it?
7. Howard: No, it gets worse . or it just complicates things as you
8 go on further down the road.
9. Judy: Can you look at your own life, kind of on a continuum?
10. Look down the road of that line and see what that's
11. gonna do . in your own life?
12. Howard: S- look on [down the road?]
13. Judy: [Yeah] Kinda visualize what um . your own
14. life will be like if you- you don't deal with some of it
15. ( ) your problems (1.5) Can you see how it might
16. complicate . [your life?]
17. Howard: [It will] just continue the way it is.

18. Judy: Kind of like *a snowball?* .[effect]
19. Howard: [No, no] *not a snowball.* Just
20.   kinda *floating, floating down the river.*
21. Judy: *Floating down the river.*
22. Howard: That's what I'm doing now. That's what I was afraid I
23.   was gonna go back into after this. I said something
24.   the first time I talked to you about
25.   Yeah
26.   *floating*
27.   and being afraid of going back into *floating.* That's
28.   just, you know, *floating, drifting..*
29. Judy: So you're *adrift* right now?
30. Howard: Yeah. And I feel dead and I feel and I'm-I drink to
31.   feel a little bit deader. No, that's not true.
32. Judy: Feel depressed (.) or numb?
33. Howard: Yeah.
34. Judy: Numb, you feel?
35. Howard: Yeah. . Yeah.
36. Judy: What's it like to be *floating down the river?* Tell me
37.   more.
38. Howard: (2) It's (1) comfortable. It's safe. . Everything just
39.   *keeps on an even keel,* you know.
40. Judy: Mmhmm.
41. Howard: You're just kinda *floating..*
42. Judy: Kind of *in a canoe? . going down the river,* or-
43. Howard: No, more like *a great ole big barge . on a great old*
44.   *big* [*river.*]
45. Judy: [*barge*] very stable, kinda.
46. Howard: Yeah. *Plenty of room to spread out* and . *sit in the*
47.   *sun.* Yeah, and you don't have to worry about
48.   *falling off the edge.*
49. Judy: Mmhmm
50. Howard: And sun, you know, it's *kinda hazy.* It's *not really*
51.   *clear sun. It's kinda hazy.*
52. Judy: Mmhmm.
53. Howard: Kinda *half asleep,* that's what it's like.
54. Judy: What happens when you kind of come to *the* (.) *falls,*
55.   *the falls that are down there, about two miles*
56.   [*down the river?*]
57. Howard: [Get the hell off] *the river*
58. Judy: That's certainly one way to handle it. Get out.
59. Howard: I feel a lot of discomfort. That's what happened last
60.   month. *I hit those falls* last month.
              ((noise))
61. Judy: I don't know why it did that. So that's what happened
62.   . um this . last time there was kind of um . an external
63.   . situation that sort of *forced you out of your boat..*
64. Howard: It was uncomfortable but I was, I was pretty, I was
65.   enjoying it too. And I didn't want to go back to just
66.   *floating.* It was uncomfortable and I was (1) out, I
67.   don't, I been *floating* a long time.
68. Judy: Mmhmm . Well, you've found that it works for you .
69.   in a sense.
70. Howard: What works for me?
71. Judy: *Floating.*

72. Howard:	Because I'm . stay . comfortable and
73. Judy:	In a sense, but it may . now be inappropriate. It may
74.	not be working as well . as it did in the past.
75. Howard:	Mm. . Yeah, I'd like to have a little excitement now
76.	and then.
77. Judy:	*Some rapids.*
78. Howard:	Yeah. ((laugh)) Something I can s- keep in control of
79.	maybe and *not drown.* But . yeah, I think I am
80.	bored.

The discourse begins prosaically, but when the client uses a dead metaphor 'down the road,' in line 8, the therapist Judy seizes upon the client's own words. She picks up *road* as signifying the future, and asks, 'Can you visualize your future, look down the road?' In line 12 the client has not yet shifted into figurative language. He questions what it is he is supposed to do: 'Look down the road?' In line 13 the therapist specifically asks him to visualize, but in line 17 he makes a prosaic response that if he doesn't deal with his problems his life will just continue in the same manner. In line 18 the therapist offers an alternative interpretation. Using a conventional metaphor, *snowball*, she asserts that, if left unattended, his problems will increase in size until they are gargantuan. So far their discourse is ordinary.

However, in lines 19-20 the client rejects the formulation, 'No, no, *not a snowball*,' and prefers to create his own metaphor. The metaphor he offers is, *'Just kinda floating, floating down the river'*. In line 22 the client has a flash of insight. He claims the metaphor, *'floating down the river*,' is pertinent to describe his present situation. 'That's what I'm doing now.' It is notable that by line 21 the client has used the words *floating* or *drifting* six times.

In line 29 the therapist ratifies his metaphor with a semantically related term, the synonym *adrift*. I draw attention to line 36, in which the therapist makes an open request for more information, using the client's metaphor and asking him to describe what *floating down the river* is like. There is a significant pause of two seconds while the client formulates a response. In line 39 he uses the phrase, *on an even keel*. This boating term allows for collocational cohesion with the *river* metaphor and is the first indication that the metaphor may be extended. At this point we have the start of a language game. An invitation to speech play has been extended and accepted.

The game continues. In line 42 the therapist broaches an extension, 'Kind of *in a canoe*?' Although the client rejects the specific choice of a boat offered, he indicates his willingness to continue playing with language by offering his preference for a type of boat. He seizes upon *'barge, a great ole big barge'* in lines 43-44 to symbolize, perhaps, the unglamorous but serviceable job as an orderly in a hospital.

A very interesting point arises in line 54 when the therapist uses an entailment of *river, 'the falls*.' She asks, 'What happens when you kind of come to *the falls*?' This question appears to probe the client's understanding of his life course. His current behavior (*floating on a barge*) could prove dangerous or disatrous if there are *falls*.. In line 59, there is a significant change in the tone of voice by the client. There is a quickened tempo, an air of excitement which may correspond with a flash of understanding. The client says, 'That's what happened last month. I hit *those falls* last month.'

In line 61 the therapist explicitly confirms that she understands that *the falls* is equivalent to his being fired last month. She gives an explicit paraphrase of *the falls*, 'an external situation that...forced you out of *your boat*.' In line 75 the client states his need for a change. 'I'd like to have a little excitement now and then' He would prefer a less monotonous existence than *floating*.

The crowning achievement of this jointly constructed extended metaphor is in the therapist's two-word metaphorical interpretation of the client's literal statement (lines 75-76) in her mention (line 77) of *some rapids*. Note that both *rapids* and *falls* are co-hyponyms of *river*. Her formulation *some rapids* is apt, cohesive, surprising, and playful.

It brings both delight and satisfaction to the client. He laughs and agrees in line 78. The client accepts the therapist's metaphor ('Yeah') and further contributes to the joint construction of extended metaphor by adding to the elaborate web of extensions they have both been engaged in weaving with the metaphorical assertion that he'd like something he can keep in control of and *not drown* in line 79. Note that *drown* is an antonym of *floating*.

In summary, we find that three major metaphors, *floating, river, barge*, are chosen by the client to represent his attitude towards life and his current work. As shown in (9), between the client (C) and therapist (T), 15 extensions revolving around these three metaphors are introduced.

(9)   **Metaphorical Extensions**
       Floating   **River**

    T   C   T   C
  adrift   drifting   the falls   those falls
  floating   keep on an   some rapids
                 even keel
                not drown

                      Barge

            T   C
          canoe?   great ole big barge
          barge   plenty of room to
       forced you out   spread out
        of your boat   sit in the sun
                              don't have to worry
                                about falling off
                                the edge
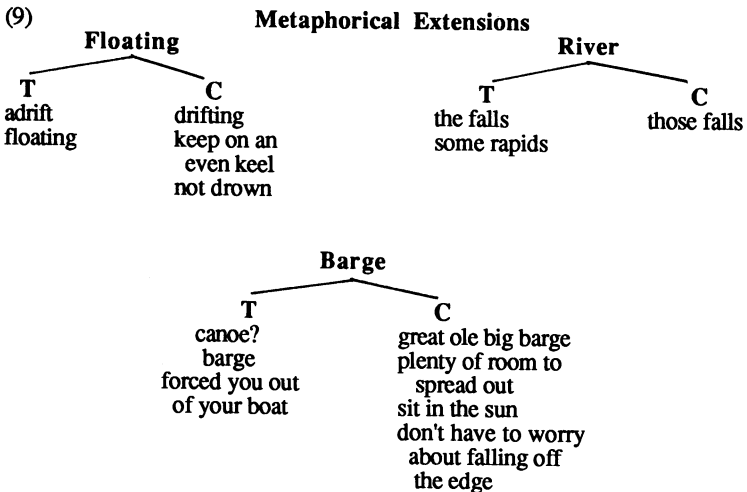
Some metaphors are accepted and some are rejected. The point is that they are all negotiated. Their meaning is not static but interactive. The discourse here is not by accident. The speech play emerges as interlocutors talk. The discourse participants have seized upon an inherent property of language, collocational cohesion, and have exploited it, magnifying the effects. In jointly accomplishing what the reader can see is a piece of verbal art, the client and therapist have experienced working together through language. They have obeyed Grice's Cooperative Principle even though they have flouted the maxims.

The samples above indicate that metaphor is a discourse strategy that enables people to talk about such troubling and important aspects of life as anger, insanity, being fired, and not living up to potential. Metaphor is not always successful, but when it is, therapists and clients find in metaphor a useful means of learning about and talking through the client's concerns in an oblique and less threatening way. I conclude that when metaphor is apt, and jointly constructed, it builds rapport between therapist and client and helps develop insight, a core goal of psychotherapy.

With its focus on dyadic verbal exchanges, this study goes beyond previous systematic observations of textual cohesion which have largely been based on written material. The study illustrates the various manners in which metaphor can be received in psychotherapy and demonstrates that collocational cohesion of metaphorical entailments can be a resource, a discourse strategy in dyadic oral interchanges. A study such as this with its emphasis on the interactional rather than static nature of metaphor advances our understanding of the many social uses of metaphor and illustrates the poetics of everyday language.

REFERENCES

Aftell, Mandy and Robin Lakoff. 1985. When talk is not cheap or how to find the right therapist when you don't know where to begin. NY: Warner Books.

Burbules, Nicolas, Gregory Schraw and Woodrow Trathen. 1989. Metaphor, idiom, and configuration. Metaphor and Symbolic Activity Journal 4(2):93-110.

Ferrara, Kathleen. (in press) Accommodation in therapy. In H. Giles, N. Coupland and J. Coupland (eds.) Accommodation in context: Directions in applied sociolinguistics. Cambridge: Cambridge University Press.

Ferrara, Kathleen. 1988. Variation in narration: Retellings in therapeutic discourse. In K. Ferrara, B. Brown, K. Walters, J. Baugh (eds.) Linguistic change and contact, pp. 100-113. Austin: University of Texas, Dept. of Linguistics.

Gordon, David. 1978. Therapeutic metaphors. Cupertino, DA: Meta Publications.

Grice, H.P. 1975. Logic and conversation. In P. Cole and J.L. Morgan (eds.) Syntax and semantics 3: Speech acts, pp. 41-58. NY: Academic Press.

Halliday, M.A.K. and Ruqaiya Hasan. 1976. Cohesion in English. NY: Longmans.

Labov, William and David Fanshel. 1977. Therapeutic discourse: Psychotherapy as conversation. NY: Academic.

Lakoff, George and Mark Johnson. 1980. Metaphors we live by. Chicago: University of Chicago Press.

Lenrow, Peter B. 1966 The use of metaphor in facilitating constructive behavior change. Psychotherapy 3(4): 145-148.

Lentine, Genine. 1988. Metaphor as cooperation in therapeutic discourse. In. K. Ferrara, B. Brown, K. Walters, J. Baugh (eds.) Linguistic change and contact, pp. 192-199. Austin: University of Texas, Dept. of Linguistics.

Martinich, A.P. 1984. Communication and reference. NY: Walter de Gruyter.

Ortony, Andrew. 1980. Some psycholinguistic aspects of metaphor. In R.P. Honeck and R.R. Hoffman (eds.) Cognition and figurative language, pp. 69-86. Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Pollio, Howard R. and Jack M. Barlow. 1975. A behavioral analysis of figurative language in psychotherapy: One session in a single case study. Language and Speech 236-254.

Pollio, Howard R., Jack M. Barlow, Harold J. Fine and Marilyn R.Pollio. (eds.) 1977. Psychology and the poetics of growth: Figurative language in psychology, psychotherapy and education. NY: Halsted Press.

Sapir, J. David & J. Christopher Crocker. 1977. The social uses of metaphor. Philadelphia: University of Pennsylvania Press.

Sherzer, Joel 1982. Play languages: With a note on ritual languages. In L. Obler and L. Menn (eds.) Exceptional language and linguistics, pp. 175-199. NY: Academic

Reflexives as Grammatical Constructions: A Case Study in Czech

Mirjam Fried
University of California, Berkeley

Reflexivization processes in Slavic and Romance languages are often described, correctly, as being similar in nature and origin (Hock, Barber, Geniušiene). They all involve the morphosyntactic form $V + Reflexive\ Morpheme$ (RM), where RM was historically a form of the reflexive pronoun. Synchronically, however, these languages differ in the range of semantic/pragmatic functions served by this form and in the syntactic behavior displayed by a given use or function. Providing a unified account of the polyfunctional distribution of RM has been the goal of many proposals within various theoretical frameworks. However, since most linguists have focused primarily on Romance and especially French data (Croft *et al.*, Wehrli, Grimshaw) their accounts offer a very limited insight into the much more fragmented system of reflexives in the Slavic languages.

Within the Slavic family, Modern Czech (MC) provides perhaps the most challenging data for establishing a coherent network of the uses of the RM. At the same time, generalizations based on a broad variety of uses can contribute more to the understanding of languages with simpler systems. The reflexive uses in Czech involve a great deal of homonymy, ranging from the straightforward anaphoric construction $V + Reflexive\ Pronoun$ (*Umyl jsem se/sebe* 'I washed myself') on the one hand to the fully lexicalized, no longer analyzable reflexive verbs (*smát se* 'to laugh') on the other, with a number of specialized but very productive constructions in between, traditionally labeled "passive", "impersonal", and "dispositional". In addition, the reflexive pronoun can appear either in the Accusative (*se*) or the Dative (*si*) form, each of which gives rise to different non-anaphoric uses.

The analysis of the anaphoric reflexives alone is a matter of some controversy, but it is especially the passive and impersonal uses that have generated the most interest - both with respect to each other and in relation to other reflexives. The situation is further complicated by the fact that the languages in question also make use of periphrastic passive forms and there are some clear distributional differences between the two that cannot be (although often have been) ignored. Focusing on two of the non-anaphoric uses of *se*, I will argue that the traditional treatments, which view reflexives primarily as either "intransitivizing" (Babby & Brecht, Grimshaw) or voice phenomena (Štícha, Barber, Croft *et al.*), cannot adequately capture their true nature. I will propose that a meaningful analysis of these uses should be based on the notion of grammatical construction (Fillmore 1988a) in which syntactic, semantic, and contextual information combine to form a uniquely identifiable linguistic unit.

Let us begin with a brief description of the anaphoric reflexive construction (AR) which is the source of the reflexive form and thus essential in any attempt to motivate the non-anaphoric uses. Without going into much detail, AR can be described as having two crucial properties: (i) a special configuration of grammatical relations, whereby the Subject marks both the Actor and Endpoint[1] arguments, and (ii) a special semantic relationship that holds between the Actor and the Endpoint and which consists in a single referent playing two semantic roles. In addition, AR has the following features that will be relevant in the subsequent discussion: AR relates an action, it does not allow one-place predicates, and the verb has active morphology.

The syntactic (as opposed to lexicalized) non-anaphoric uses of *se* have been generally treated as some sort of passive, to greater or lesser degree identified with the periphrastic passive (Kopečný, Parolková, Comrie, Barber, Grimshaw). Possibly with the exception of Barber's analysis, this view is largely motivated by their superficial resemblance to the passive in that the Endpoint is promoted to the hierarchically more prominent position of Subject. While this criterion can indeed be part of their description, it is by no means sufficient for fully capturing the character and function of these constructions. I will base my analysis on the hypothesis that the special configuration of syntactic relations characteristic for AR was generalized (hence the similarity with the passive) and that it turned *se* into a grammatical marker with precisely that function - to signal a marked hierarchy of constituents. However, this syntactic function is accompanied by special semantic and pragmatic features (some of them shared with other, non-passive constructions) that make these uses of *se* markedly distinct from the passive.

As noted above, the special configuration in AR can be formulated in terms of two conditions that are fulfilled simultaneously: the Subject marks the Actor, and at the same time, it marks the Endpoint. The former relates AR to the plain active configuration, the latter relates it to passivization. Unlike passive, though, the reflexive constructions do not involve passive morphology on the verb, and the Actor is demoted to the point of complete suppression. Because of this difference these uses of RM have been labeled "mediopassive", "middle", "agentless passive", or (also in the Czech linguistic tradition) "reflexive passive". There seems to be no *a priori* reason to reject this "passive" analysis for sentences exemplified by (1) where the Endpoint participant is instantiated as the Subject, the verb remains in the active form, and the Actor is unexpressed:

(1) Bude   se tu stavĕt   nemocnice.
     be-AUX-3SG-FUT RM here build-INF hospital-NOM-SG
     'Some hospital-building will be going on here.'

The problem is that Czech has similar looking constructions built around one-place predicators (2):

(2) Šlo   se tam   dlouho.
     walk-3SG-PAST-N(eut.) RM there long-ADV
     'It took a long time to walk there.'

The sentences exemplified by (2) have usually been called "impersonal reflexive" and treated as clearly distinct from "reflexive passive" (e.g., Trávníček; Bauer & Grepl). Elaborating especially on Kopečný's proposal (1954), I will try to show that there are more similarities than differences between them and that they could be analyzed as instances of a single Generic-Actor Reflexive construction (GR) with its own distinct set of properties. In clarifying their relationship to passivization, I will examine the role of Object promotion vs. Subject demotion, the relationship of GR to other generic-agent constructions, the distribution of aspect, and the nature of the Actor in this construction.

First, let us look at the passive vs. impersonal character of the two traditionally recognized types that is often cited as a major reason for separating them. Schemati-

cally, they can be represented as (4) and (5); for easier reference, I also include the periphrastic passive (3):

(3) $\text{NP}_{\text{NOM,Pat}}$ - Vfin_{be} - V_{PASS} - $(\text{NP}_{\text{INST,Agt}})$ = Periphrastic Passive
(4) $\text{NP}_{\text{NOM,Pat}}$ - $\text{Vfin}_{\text{ACT(ive)}}$ - *se*   = Reflexive Passive
(5) $\text{Vfin}_{\text{3Sg.NeutACT}}$ - *se*   = Impersonal Reflexive

The traditional argument in support of this tripartite division goes as follows: the periphrastic passive and reflexive passive should be grouped together because they both promote the Endpoint of a transitive event into the Subject position, whereas the impersonal reflexive is reserved for verbs without an Endpoint argument. This purely syntactic treatment implies that (i) both passive constructions are limited to transitive predicators only and (ii) the Endpoint is always expressed (as Subject).

There are, of course, several problems with this division. As Kopečný points out (p. 226), it has, for example, nothing to say about sentences of the type (6):

(6) Bylo   tu   zneužito   naší   důvěry.
    be-AUX-3SG-PAST-N here   abuse-PASS-N   our   trust-GEN
    'Our trust has been violated.'

(6) is an example of a perfectly well-formed instance of periphrastic passive morphology formed from an intransitive verb, thus excluding even the possibility of Object promotion as the primary measure of its passivity. This sentence also has its (practically synonymous) impersonal reflexive counterpart (6'):

(6') Zneužilo   se   tu   naší   důvěry
     abuse-3SG-PAST-N RM here our   trust-GEN

It seems hard to justify that (6') should be analyzed as unrelated to (6) (i.e., impersonal reflexives vs. periphrastic passives), especially if (7) and (7') are not so analyzed (they are both considered "passive"):

(7) Hrad   bude   opraven   za pět   let.
    castle-NOM-SG be-AUX-3SG-FUT repair-PASS   in five   year-GEN-PL
    'The castle will be renovated in five years.'

(7') Hrad   se   opraví   za pět let.
     castle-NOM-SG RM repair-3SG-PRES in five year-GEN-PL
     'The renovation of the castle will take five years.'

This, however, ceases to be a real problem if we abandon the formal signs of transitivity as the crucial criterion and instead focus on the underlying argument structure. The second argument of *zneužít* 'abuse' is encoded as Genitive, which prevents it from promotion to Subject in Czech, but it still is a two-place predicate with the general structure "Actor-Endpoint". This means that the impersonal form in (6) is just a formal requirement imposed by a specific verb rather than a signal that the impersonal form reports a state of affairs inherently different from that associated with personal passive.

More damaging (and apparently not challenged by Kopečný) is the fact that, by following the principle of Object promotion as the crucial unifying property of passives, the traditional account is equally inconsistent in subsuming under the label "passive" constructions that are morphosyntactically heterogeneous (periphrastic passive and reflexive active) and furthermore, in separating passive reflexives in which the Endpoint argument is overtly expressed from those which leave it out under the Indefinite Null Complement interpretation (Fillmore 1988b:94). Thus (8a) would be classified as an instance of passive reflexive, while (8b) would be an instance of an active impersonal reflexive construction, even though they both involve a clearly transitive verb:

(8) a. Večeře   se jedla   v   sedm.
       dinner-NOM-SG-F RM eat-3SG-PAST-F at   seven
       'They (generic) ate dinner at seven o'clock.'

    b. Jedlo   se v   sedm.
       eat-3SG-PAST-N RM at   seven
       'They (generic) ate at seven.'

The "impersonal reflexive" is thus characterized in contrast to "passive" as a construction involving only Subject deletion. This, of course, is a classification based strictly on surface syntactic forms and ignores not only the distribution of the underlying semantic roles (thus remaining incapable of explaining which verbs can or cannot participate in each of these constructions) but also the pragmatic content of each construction, which has to be taken as one of the distinctive parameters in contrasting passives vs. reflexives.

The question of what are the crucial processes involved in passivization crosslinguistically has been very insightfully investigated by Comrie in his attempt to provide a unified account of personal and impersonal passives (1977). His general line of argument is very appealing, particularly so for classifying sentences of the kind illustrated in (9), which stand half way between a promotional passive (the Object is promoted to Subject) and an impersonal construction (the verb does not agree with the Subject but retains its impersonal form 3sg,Neuter):

(9) Zebralo   sa smetana.   (dial.)[2]
    skim-3SG-PAST-N RM cream-NOM-SG-F
    'One skimmed the cream.'

To overcome this problem, Comrie proposes "Subject Removal" as the link between both kinds of passive rather than contrasting them around the property of "Object Promotion". Among the languages he examines is also Polish, which does not require Object promotion as part of passivization, allowing the configuration exemplified by (9) above in both the reflexive and the periphrastic passive sentences (p. 49). However, by treating both forms as morphosyntactic variants of the semantic category passive, Comrie ignores some significant differences associated with each form. First of all, one-place predicates do not undergo passivization (cf. Czech *bylo jděno 'it was walked'), but they can freely occur in the reflexive construction (2). Secondly, there is the question to what extent the so-called passive reflexives (1, 7, 8) are really passive

semantically and pragmatically. And finally, periphrastic passive has more of a stative flavor, whereas the reflexive construction clearly expresses an action or a process; this difference is manifested by the patterns in the distribution of aspect (perfective/imperfective). The latter two issues will be taken up shortly in more detail.

In light of these facts, I propose that these reflexive constructions could and should be classified in constrast to periphrastic passive rather than as partly coinciding with it and that both passive and GR constructions should be viewed as falling into two categories: personal and impersonal. The distinction between personal and impersonal form lies with the valence properties of a given predicator, not with the category passive vs. reflexive. The traditional grouping of these forms (10a) pits the category passive against impersonal; my proposal (10b) recognizes two distinct levels: on the formal level, the personal form contrasts with impersonal, while on the "constructional" level the passive construction constrasts with the GR construction:

(10) a. **Passive** ⟨ Personal / Impersonal ⟩   vs.   **Reflexive** ⟨ Passive / Impersonal ⟩

b. **Passive** ⟨ Personal / Impersonal ⟩   vs.   **GR** ⟨ Personal / Impersonal ⟩
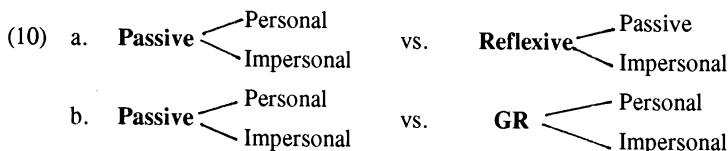
As noted above, Comrie's proposal that Object promotion be only secondary to Subject demotion in determining what does or does not qualify as "passive" ignores the difference between periphrastic passive and reflexive constructions. Nevertheless, if we separate the two forms on the grounds that they manifest different semantic and pragmatic behavior, his solution can still be successfully applied to handle both the true passives (periphrastic) on the one hand and Generic-Actor reflexives on the other.

We can summarize, then, that both the "passive reflexive" and the "impersonal reflexive" obligatorily suppress the Actor argument and whether or not there is an overt Subject is determined by the predicator rather than by the construction, according to the following language-specific principle (here stated for standard MC):

(11)   If the predicator requires a Direct Object, it is realized as Subject, which in turn determines the agreement on the verb.

This analysis offers a further advantage in that it can also eliminate the exceptional status of the marginally occurring reflexive sentences exemplified in (12), where the Subjects are instantiated by a non-3rd person:

(12) a.   Počítám   se   mezi   jeho   stoupence.   - ambiguous
          count-1SG-PRES   RM   among   his   supporter-ACC-PL
          'People consider me to be one of his supporters.'
          'I consider myself one of his supporters.'

b.   Vyzýváte   se,   abyste   (dlužnou částku splatil do tří dnů.)
          request-2PL-PRES   RM   CONJ-PURPOSE
          'It is requested of you that (you pay your debt within three days.)'

Kopečný treats these forms as non-systemic (p. 229), but under the present analysis, their unexpected verb forms may be just natural consequences of the above-mentioned principle governing the realization of Subject in this construction: if the Direct Object is instantiated by the first or second person pronoun, the verb form has to change accordingly. There is no need to make separate statements (or rules) to account for these seemingly anomalous sentences.

In addressing the question of the passive meaning of GR, we of course can first note that the verb has active, not passive morphology. But more importantly, we can establish their relationship to other generic-agent constructions in the language that certainly are not passive. Kopečný notes (p. 227) that other Slavic languages offer good evidence of the "active" character of GR because the passive reflexive constructions in Czech often have non-reflexive active equivalents in other Slavic languages - compare the following examples from Czech (13a) and Russian (13b):

(13) 'Coal is mined here.'
  a. zde   se   dobývá   uhlí   = reflexive passive
      here   RM mine-3SG-PRES   coal-NOM-SG

  b. zd'es'   dobivayut   ugol'   = non-reflexive active
      here   mine-3PL-PRES   coal-ACC-SG

But in fact, MC itself exhibits similar alternations where the reflexive passive is more naturally paraphrased by a non-reflexive active sentence than by periphrastic passive. Czech grammarians usually describe the class of (13b) sentences as "generic-subject sentences", very commonly used in contexts where the Actor is not individually identifiable (Panevová (1973) refers to them more appropriately as "generic-agent" (p. 133)). In these non-reflexive sentences, the generic nature of the Actor is marked by the third person plural form of the verb with the surface Subject NP obligatorily absent. Czech usually reserves this construction for verbs of announcing and *verba dicendi* in general such as *říkat, povídat* 'say', *hlásit* 'announce', *ukazovat (v televizi)* 'show (on TV)', *telefonovat, volat* 'telephone, call', etc., although it is not limited to them exclusively. They also have some special properties of their own which distinguish them from GR, such as the exclusion of the speaker from the collective Actor (Panevová:139), but their pragmatic closeness to the GR construction is obvious. In either class, the absence of a surface Actor does not in itself constitute passive meaning; it only signals an extreme indefiniteness of the Actor.

Another characteristic associated with GR that sets it apart from the passive is the well-known fact that GR sentences often have a modal interpretation, especially in the present tense. They are used as general statements of advice, orders, reprimands, etc., and even the personal variety (14a) cannot be interpreted as passive:

(14) a. Jedí   se jen   ta   zrníčka.
      eat-3PL-PRES RM only   that-NOM-PL   seed-NOM-PL
      'You (generic) can only eat the seeds.' (in a pomegranate, for example)

  b. U   nás se klepe   na dveře, než   se   vejde.
      with us RM knock-3SG-PRES on door   before RM   enter-3SG-PRES
      'You (generic) first have/are supposed to knock on the door before you enter.'

It seems, then, that the active-passive distinction does not offer a very reliable dividing line between the "reflexive passive" and "impersonal reflexive", not only for formal reasons (both share active morphology on the verb), but especially because they fit the same pragmatic/semantic description distinctly different from the periphrastic passive.

Another possible argument for rejecting the grouping of the reflexive passive with the periphrastic passive (rather than with the impersonal reflexive) is the category of aspect. It is generally the case that the periphrastic passive has a resultative flavor and shows preference for the perfective aspect (15a) while the reflexive passive strongly favors imperfectives (15b):

(15)   'The castle was founded in 1348.'
a. Hrad   byl   *zakládán/založen   v roce 1348.
   castle-NOM-SG   be-AUX-3SG-PAST   found-PASS-IPF/PF   in year

b. Hrad   se   zakládal/*založil   v roce 1348.
   castle-NOM-SG   RM   found-3SG-PAST-IPF/PF   in year

Let us now briefly address the question of the Actor argument in these construc-tions. It is an accepted fact that the GR construction presupposes a human Actor, which is interpreted as generic or anonymous[3]. Given this "default" quality of the Ac-tor we might expect an Agent as the most likely semantic role filling the Actor argu-ment slot. The following data will show whether this prediction is correct:

(16) a.   *Slyšel   se   hluk.
           hear-3SG-PAST RM   noise-NOM-SG
           'Some noise was heard.'

      b.   Poslouchala   se   hudba.
           listen-3SG-PAST RM   music-NOM-SG
           'One listened to music.'

      c.   Žilo   se   velmi   skromně.
           live-3SG-PAST-N RM   very   modestly
           'One lived very modestly.'

Apparently, verbs with the Experiencer-Stimulus (rather than Agent-Patient) frame do not fit in this construction (16a). As as result, verbs of perception in general are excluded (slyšet 'hear', vidět 'see', cítit 'feel', etc.) when used in their literal sense. This restriction does not apply, predictably, to the metaphorical uses of these verbs:

(17) a.   To se hned   tak nevidí.
           it RM immediately so   NEG-see-3SG-PRES
           'You (generic) don't run into something like this very often.'

      b.   To se ještě uvidí!
           it RM still   see-3SG-PRES
           'We'll see about that!'

We should note, however, that both of these examples also have to be regarded as fixed expressions with the structure of GR rather than fully productive instances of GR (they only occur in the present tense indicative assertions and, in general, their use is highly idiomatic).

(16c) might seem slightly more problematic if we maintain that verbs like *žít* 'live' have a Patient as their single argument (Fillmore 1988b:68). To allow that the Actor in GR can be either an Agent or a Patient would mean that we then have to stipulate what kind of a Patient it must be in order to capture the fact that this Patient corresponds to a very different position in the semantic role hierarchy from the one associated with the Patient of a typical transitive event. In other words, we would need some special mechanism for distinguishing this Patient from the typical Patients that regularly occupy the Subject position in this construction. However, Fillmore admits that at least some of these Patient-Subject predicators do exhibit somewhat different behavior from truly Patient-Subject verbs such as 'break', 'open', etc. (ibid.:81). It seems that this semantic role would be best described as an Undergoer of the action. For the lack of a better term I am borrowing this label from the Czech linguistic tradition as a loose translation of the term "nositel děje". This term is used to describe an entity that can appear in the Subject argument slot of one-place predicates but that is neither a typical Agent (i.e., an initiator) nor a typical Patient of a transitive event. For example, the single argument of such predicators as *umřít* 'die', *žít* 'live', *spát* 'sleep', etc., is regarded as "the undergoer of an inherent action (*'nositel inherentního děje'*) which is characteristic of or fundamentally connected with the existence or functioning of a given entity" (MČ III:46). We can then generalize that the Actor in GR is either an Agent or an Undergoer in the sense just described. Needless to say, the semantic constraint on the Actor is of course a characteristic property of GR and does not apply to the periphrastic passive - yet another property that sets the two constructions apart.

By appealing to this constraint we can also explain the different acceptability judgements in sentences (18-19) below, which look structurally very similar to each other:

(18) *Hrad   se zakládal   Karlem IV.
      castle-NOM-SG RM found-3SG-PAST Charles-INS
      'The castle was founded by Charles IV.'

(19) Prach   se pohlcuje   filtrem.
      dust-NOM-SG   RM absorb-3SG-PRES filter-INS-SG
      'Dust is absorbed by a filter.'

(18) is unacceptable because it violates the requirement that Actor be generic; (19), on the other hand, is grammatical because the Instrumental NP marks an Instrument, not a volitional Agent or a human Undergoer.

The sentence in (19) invites comparison to similar-looking sentences such as in (20), and it has been suggested that (19) and (20) are instances of the same phenomenon (Štícha:190):

(20) Dveře   se   zavřely   (průvanem).
    door-NOM-PL   RM   close-3PL-PAST   wind-INS-SG
    'The door closed (in the wind).' (lit. 'by the wind')

Morphosyntactically, both (19) and (20) show the same pattern: NP_{NOM} - se - Vfin. - NP_{INS}. Also, se is clearly not anaphoric in either instance. However, while in (19) INS encodes an Instrument, which requires the presence of an Agent, in (20) it encodes a physical force, which precludes another agentive participant. This claim can be tested by the following sentences:

(19') Prach   pohlcujeme   filtrem.
     dust-ACC-SG   absorb-1PL-PRES filter-INS-SG
     'You (generic) can absorb the dust by a filter.'

(20') *Dveře   zavíráme   průvanem.
     door-ACC-PL close-1PL-PRES   wind-INS-SG
     'You (generic) can close the door by wind.'

(19) and (20) obviously express two different kinds of events: we can interpret the door in (20) as capable of acting upon itself under certain conditions, but we certainly cannot ascribe this quality to dust absorbing. Nevertheless, some relationship between the two uses undeniably exists and should be dealt with.

Part of the problem with Štícha's treatment (which is in many ways very inspirational) stems from his commitment to handle the non-anaphoric functions of RM in terms of voice. He proposes a three-way distinction of voice: active, passive, and reflexive ("intransitive" in his terminology). The function of the latter is said to report an "autonomous process implying two semantic roles" (p. 189). While this strategy works well in accounting for the behavior of (20), it forces him to treat both (20) and (19) indiscriminately as "intransitive" voice although they display a different degree of conventionalization (in (19) se is not integrated into the meaning of the verb the way it is in (20)), and there is nothing spontaneous about absorbing dust in (19) the way door-closing can be.

I suggest that the two sentences represent two different processes involving the weakening of the anaphoric se: lexicalization and grammaticalization, respectively. In both cases the cause of the action (Agent) is placed outside the syntactically most prominent position of Subject, which in turn weakens the syntactic status of se because the requirement of the reflexive pronoun to have an antecedent is no longer fulfilled. It follows that both cases also exclude predicators with the Experiencer-Stimulus frame (there are no reflexive verbs *vidět se, *slyšet se, etc.). But it seems to be the nature of the cause that separates the two processes: if the cause can be some independent medium (i.e., the entity expressed by the Subject NP is understood as capable of acting upon itself, possibly through some catalyst force), the Endpoint argument in the Subject position "absorbs" the role of the Agent. Since the only participant left is an inanimate object, the implicit self-affecting of the Subject, brought about by the reflexive form, is interpreted as accidental (or spontaneous) and as such is incorporated into the meaning of the verb as a marker of an unanalyzed, one-participant "autonomous process". This kind of weakening then results in the derivational function of se, giving rise to one-place reflexive predicators such as zavřít se 'close', rozbít se 'break',

naplnit se 'fill', etc.) with the argument structure Undergoer - (Cause); the parentheses indicate optionality.

This analysis can also be extended to the reflexive verbs whose only participant is an animate entity, but its self-affecting is/can be construed as unintentional (*zabít se* 'kill oneself', *vzbudit se* 'wake up', *utopit se* 'drown accidentally', etc.). In these cases, the distinction between Actor and Endpoint is significantly blurred, the cause of the action is placed outside of the Subject's volitional control, and the semantic role associated with this argument can thus be again described as an Undergoer. In other words, by collapsing the Agent and Patient arguments into one semantic role which does not imply volition (Undergoer), the same process can apply to verbs with both inanimate and animate Subjects, converging them around the property of "unintentional action". Diagrammatically, the shifts can be represented as follows:

(21) *zavřít*   -->   *zavřít se*   *zabít*   -->   *zabít se*

   Agent Patient   Undergoer Cause   Agent Patient   Undergoer Cause
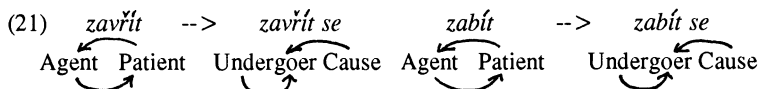
If, on the other hand, the Subject NP cannot be interpreted as capable of bringing an action onto itself, a human Agent/Undergoer is presupposed as the default immediate cause of the action, and the sentence is still interpreted as a two-participant event (1, 7', 8a, etc.) with *se* becoming a grammatical marker of the Actor's subordinate role. The grammaticalization obviously had to start with two-place predicators of the Agent-Patient variety, but once the pattern is established there is nothing in the GR construction *per se* that would prevent inherently one-place predicators from entering this construction as well (possibly with some added support from the verbs allowing an Indefinite Null Complement for their Direct Object), as long as their frame contains a human Actor.

To some extent, this division may also reflect the difference in the semantic requirements a predicator imposes on its Actor argument. The lexicalization tends to go with verbs that can have a physical force as the Actor (*zavřít, otevřít, rozbít, zabít, vzbudit*, etc. although *naplnit* and *utopit* do not quite fit this description), whereas the grammaticalization process takes place exclusively with verbs that require an animate Agent/Undergoer as their Actors.

Since it is our understanding of the world (the experience we have with door-closing or window-breaking as compared to the experience of building houses, eating dinners, or absorbing dust) that plays a crucial role in distinguishing between these two functions of *se*, it is obvious that a strictly formal analysis cannot adequately deal with them. The surface similarities in fact give rise to homophonous uses (illustrated in (22); the data is partially quoted from MČ II:176) that can be differentiated only through the pragmatics. The above analysis seems to offer a principled way of applying the pragmatic criterion:

(22) a. Zabil   se z   nešťastné   lásky.   - anaphoric *se*
     kill-3SG-PAST-M RM from   unhappy-GEN love-GEN
     'He killed himself because of unhappy love.'

    b. Zabil   se pádem   ze střechy.   - derivational *se*
     kill-3SG-PAST-M RM fall-INS   from roof-GEN
     'He got killed by falling from a roof.'

c. Na svátky   se zabila   husa.   - grammatical *se*
   for holiday-ACC-PL RM kill-3SG-PAST-F goose-NOM-SG
   'One killed a goose for the holiday.'

The preceding discussion should now provide ample justification for collapsing (4) and (5) into one Generic-Actor Reflexive construction (the abbreviation 'gn' stands for 'generic-null' and refers to the null realization of the NP on the surface):

(23) $(NP_{NOM,Pat})$ - $Vfin_{ACT}$ - *se* - $NP_{gn,Actor}$

with the following properties:

(24) - *se* is a grammatical morpheme indicating that Subject, if present, does not mark the Actor;
      - GR is indifferent to transitivity, i.e., allows both transitive and intransitive predicators;
      - the verb has active morphology; the Subject-verb agreement is determined by the valence requirements of the verb;
      - preference for the imperfective aspect;
      - Actor is an Agent or human Undergoer;
      - Actor is presupposed as either generic or anonymous;
      - GR relates an action/process.

Now that we have identified the fundamental characteristics of GR, we can attempt to summarize its relationship to the Anaphoric Reflexive construction, as well as to the periphrastic passive. The GR construction is linked to AR through three properties: (i) it maintains active verb morphology (unlike passive), (ii) it shows a marked hierarchy of constituents (like passive), and (iii) it relates an action. With respect to the latter, GR then appears as a non-stative counterpart to the passive in that it is also an Agent-demoting construction. Unlike passive, however, the center of attention is on the action itself, not on the object or result of the action.

The analysis presented in this study demonstrates that it is possible to treat functionally distinct reflexive forms as mutually related (with a smaller or greater degree of transparency, of course) if we take into consideration more than just their surface representation. If our analysis includes semantic and pragmatic properties of individual forms as equal partners in building up a construction in Fillmore's sense, relationships that are syntactically quite diverse may converge along the other two parameters. This treatment carries several advantages: in addition to allowing clear distinction between functionally and formally diverse forms (reflexive vs. periphrastic passive), it also offers a principled way of separating homonymous uses (e.g., Anaphoric vs. Generic-Actor vs. lexicalized reflexives); it avoids the necessity of forcing a binary choice between a syntactic analysis and full lexicalization in those instances where neither would be appropriate, thus allowing more natural inclusion of idiomatic expressions; and finally, it might bring some more insight into the complex lexicalization patterns associated with the Czech (and Slavic) reflexives.

Notes:
1. Since the arguments underlying the Subject and Object positions can be associated with a variety of semantic roles, I refer to them by the superordinate categories Actor and Endpoint, respectively.
2. The data is quoted from the field work on East-Moravian dialects (Chloupek & Šlosar:236).
3. The Czech syntacticians usually label these constructions "deagentive" to indicate the extreme indefiniteness of the Actor (Bauer & Grepl, MC, *inter alia*), and the term "agentless passive" has been proposed for French (Grimshaw). I find both terms misleading because they suggest that there is no Actor (or Agent) to consider. I also believe that it cannot be dismissed as just a terminological issue. Only if we make the presence of the Actor (albeit not manifested on the surface) a cornerstone of our analysis can we fully predict which verbs can be accommodated by this construction.

Bibliography:
Babby, L.H. & R.D.Brecht (1975), The Syntax of Voice in Russian, Lg 51:342-367.
Barber, E.J.W. (1974), Voice - Beyond the Passive, BLS 1:16-24.
Bauer, J. & M.Grepl (1972), Skladba spisovné češtiny (Syntax of Literary Czech), Praha: SPN.
Chloupek, J. & D.Šlosar (1972), Specifický východomoravský typ deagentních vět (A Specific East-Moravian Type of Deagentive Sentences), Studia Slovaca 7:236-239.
Comrie, B. (1977), In Defense of Spontaneous Demotion: The Impersonal Passive, in Cole P. & J.M.Sadock (eds.), Grammatical Relations, Syntax and Semantics 8:47-58, New York: Academic Press.
Croft, W. & H.B.Shyldkrot, S.Kemmer (1985), Diachronic Semantic Processes in the Middle Voice, ms.
Dušková, L. (1976), Reflexivity in Czech and in English, Studie z moderní filologie II:5-89.
Fillmore, C.J. (1988a), Mechanisms of "Construction Grammar", BLS 14:35-55.
Fillmore, C.J. (1988b), Construction Grammar, ms.
Geniušiene, E. (1987), The Typology of Reflexives, New York: Mouton de Gruyter.
Grimshaw, J. (1982), On the Lexical Representation of Romance Reflexive Clitics, in J. Bresnan (ed.), The Mental Representation of Grammatical Relations:87-148, MIT Press.
Hock, H.H. (1988), Principles of Historical Linguistics, Berlin: Mouton de Gruyter.
Kopečný, F. (1954), Pasívum, reflexívní forma slovesná a reflexívní sloveso (The Passive, the Reflexive Form, and the Reflexive Verb), in Studie a práce linguistické I, Praha: Academia.
MČ: (1988) Mluvnice češtiny II, Inflectional Morphology, Praha: Academia.
(1988) Mluvnice češtiny III, Syntax, Praha: Academia.
Panevová, J. (1973), Věty se všeobecným konatelem (Sentences with the Generic Agent), in Studia Slavica Pragensia:133-144.
Parolková, O. (1967), K problematice zvratných sloves a tzv. zvratného pasíva v současné spisovné ruštině a češtině (Reflexive Verbs and the So-called Reflexive Passive in Contemporary Literary Russian and Czech), Slavia 36:33-46.
Štícha, F. (1981), K syntakticko-sémantické konkurenci aktívních, participiálních a reflexívních konstrukcí (On the syntactic and semantic properties of the active, participial, and reflexive constructions), Slovo a Slovesnost 42:177-182.

Trávníček, F. (1956), Historická mluvnice česká - Skladba (The Historical Grammar of Czech - Syntax), Praha: SPN.

Wehrli, E. (1986), On Some Properties of French Clitic *se*, Syntax and Semantics 19:263-283.

The Role of Trigger-Target Similarity in the Vowel Harmony Process

Mary Hare
University of California, San Diego

I. Introduction[1]

The current paper proposes a connectionist processing account of certain aspects of vowel harmony in Hungarian. The paper has two interrelated goals. First, it offers an explanatory account of the behavior of the so-called *transparent* vowels in that language. Second, this account relies crucially on a connectionist theory of sequential processes: thus, to the extent it succeeds, it demonstrates the utility of connectionist models as explanatory tools in the study of linguistic phenomena.

The paper is organized in the following manner: I first review some facts about the vowel harmony process in Hungarian which present difficulties to analysis. Second, I introduce the model of sequential processes developed by Jordan (1986). The core of the paper then involves a series of parametric studies, whose aim is to determine the conditions on assimilation in a network of this type. Having established what factors constrain assimilation in the sequential network, I return to the Hungarian data and show that the interaction of these same factors predicts the correct pattern of behavior for both harmonic and transparent vowels in that language.

II. Hungarian vowel harmony (I)

The Hungarian vowel system is as shown below. This is a seven-vowel system, and each vowel has a long counterpart which is phonemic. Notice that there is a round - nonround distinction among the non-low vowels, and that while /e:/ is a mid vowel, /e/ is analyzed as being low.

	front:		back:
	unrounded:	rounded:	
i	i:	ü ü:	u u:
	e:	ö ö:	o o:
	e		a a:

Hungarian exhibits front-back harmony: in general roots contain only front or only back vowels, and suffix vowels alternate to agree in backness with those of the root. The following data, taken from Vago (1976), exemplify the harmony phenomenon. These are examples of consistent front- and back-vowel roots, followed by the dative suffix. Note that after a front root the suffix takes the form *nek*, while after a back root the same suffix is realized as *nak*.

(1) Front roots:

iker 'twin'	iker-nek 'twin-DAT'
tükör 'mirror'	tükör-nek 'mirror-DAT'

(2) Back roots:

	varos 'city'		varos-nak	'city-DAT'
	kapu 'gate'		kapu-nak	'gate-DAT'

There are a number of exceptions to the pattern of consistent harmony within roots. The most important exception, which I will concentrate on in this paper, involves the class of non-low front unrounded vowels (/i/, /i:/, /e:/, and on some accounts /e/). As the following examples demonstrate, these can also appear in the same root as a back vowel, and if they do, the back vowel of the root determines the backness quality of the suffix vowel, while the front vowel is ignored. This set of front vowels is commonly referred to as being *transparent* to the vowel harmony process, in the terminology introduced by Clements (1980).

(3)   bika 'bull'   bika-nak   'bull-DAT'
        izom 'tendon'   izom-nak   'tendon-DAT'

(4)   kocsi 'carriage' kocsi-nak 'carriage-DAT'
        taxi   taxi-nak 'taxi-DAT'

Note, however, that in certain environments these same vowels loose their transparent status, which is to say they *do* determine the backness value of the suffix vowel. This is the case if the root contains only front vowels, as in (1), or if the root ends in a sequence of such vowels, as in the (borrowed) forms shown below.[2]

(5)   aspirin   aspirin-nek
                     * aspirin-nak
        bronkitis   bronkitis-nek
                     * bronkitis-nak

The problem, then, is that an identical vowel may behave harmonically in one environment, while violating harmony in another. This complication has often been dealt with in the literature by positing a number of different sources for the segment in question, and allowing the harmonic-transparent distinction to follow from this (Clements 1980, van der Hulst 1985, Ringen 1988, among others). One drawback to such an approach is that there is no reason to establish these differences in derivational source except to distinguish between harmonic and non-harmonic behavior: there is no other behavior in the phonology of the language that motivates it. A preferable account would be one in which these differences in behavior follow from general conditions on the model. In what follows I will use a connectionist model of assimilation to suggest one such account.

III. An account of sequential process in the connectionist framework

The account that is being developed here relies on the theory of sequential processes developed by Michael Jordan. Jordan 1986 describes an interesting series of models of coarticulation effects, using a recurrent connectionist network which

learns to produce an ordered sequence of output patterns in response to a given input. The network is illustrated in Diagram I.

DIAGRAM I

PLAN   STATE
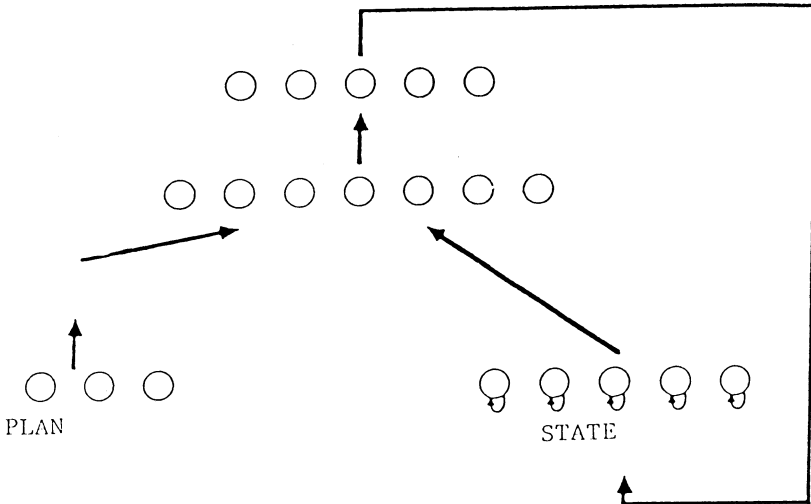
(Jordan 1986)

These models involve at least three layers of processing units: input, internal (or *hidden*), and output. Activation passes from the input to the output along weighted connections.

Input to the model consists of two parts. The first, labeled the *plan*, is an arbitrary vector that triggers the production of a given sequence. In addition, the *state* of the system (that is, the current output) is fed back over fixed connections and constitutes part of the input at the next cycle. This serves as a temporal context and aids the system in learning what part of the sequence is the next to be produced.

Learning is accomplished through the Back Propagation of Error algorithm. After each input is presented, the output that results is compared to the desired output, or *teacher*, and the discrepancy between the two is computed. This discrepancy is the *error* on that pattern. The weights on the connections are modified slightly to minimize this error. This process is repeated until some criterion of acceptability is reached.

In the simulations to be discussed here, as in Jordan's coarticulation model, output at any given time consists of a single phoneme represented by a vector corresponding to a distinctive feature description. A word or other longer sequence is represented over time as a string of phonemes on successive output cycles. Thus we can take each of the output units in the diagram above to represent a distinctive feature of a phoneme. An activation value of 1 on that unit represents [+ F], while

an activation of 0 stands for [- F]. The example below illustrates one such output layer, and the patterns of activity representing certain segments.

(6) back high low round

Examples: i = 0 1 0 0
              a = 1 0 1 0

An interesting property of this network is that particular features of a phoneme can be left unspecified for any value. The next example illustrates how such a pattern is represented in the network. In this example, the segment being described is [+ high], [- low], and [- round], but has no specification for the first feature, [back]. Hence this pattern, given here as "I", represents a segment that is ambiguous between /i/ and /ɨ/. (Lack of a specified value is indicated by the asterisk.)

(7) I = * 1 0 0

To say that a unit is "unspecified" for a value means that no error signal is propagated back from that unit. Instead of learning to match a particular teacher, or target phoneme, the feature picks up its specification from some other pattern in the sequence.

In what follows I will use the term *assimilation* to refer to the tendency of an unspecified output unit (hereafter a *don't care* unit) to take on a value influenced by one of its neighbors in time. Jordan shows that outputs tend to follow as smooth a trajectory as possible: thus a don't care unit might be expected to assimilate most strongly to its immediate temporal predecessor. In certain cases, however, the don't care unit ignores its immediate predecessor, and takes on a value close to that of an earlier pattern in the sequence. The question, then, is what factors determine the source of assimilatory influence. Note that this is exactly the problem in the Hungarian data, as well.

IV. Conditions on assimilation in the sequential network

The following set of simulations is aimed at answering this question. These simulations were designed to test the hypothesis that the similarity between vowels (in a sense which will be made precise below) is a crucial factor in determining the choice of assimilatory trigger.[3]

a. Stimuli were output sequences as in the example below.

(8)   A   0 1 0 0 0 1 0
            B   1 0 1 1 1 0 1
            C   * 1 0 0 0 1 0

Each output was a seven-bit distributed pattern, and for each plan the network learned to produce a three-pattern sequence. Members of the sequence will be referred to as A, B, and C. In each sequence the first two patterns (A and B) are

specified for all seven units, while the third (C) has one don't care unit, in initial position in the string. A and B have opposing values on this first unit.

Sets of patterns were devised in which the final two lines (B and C) were held constant with a certain number of units in common. This measure of "units in common" is referred to as the "hamming distance" between B and C and is a measure of vector similarity. The first line in the sequence (line A) was varied in similarity to the other two by manipulating the hamming distance between them. This was done in the following way. The pattern given above was the first 3-line sequence in one such set. Here B and C have opposing values on the last six bits, while A and C are identical. In the second sequence of this set (example 9), B and C remain unchanged, while A is varied to differ from C on one unit. In the third sequence (given in 10), A and C differ on two units.

(9)   A   0 0 0 0 0 1 0
         B   1 0 1 1 1 0 1
         C   * 1 0 0 0 1 0

(10)   A   0 0 1 0 0 1 0
         B   1 0 1 1 1 0 1
         C   * 1 0 0 0 1 0

This process was repeated, steadily decreasing the similarity between A and C until the set consisted of seven 3-line sequences. Note that by *similarity* I am speaking of hamming distance, a measure of overall vector similarity, and not simply the presence of similar values on any single unit.

b. Training

These sequences served as teaching output to the network described above. This network was trained on each sequence for 2000 iterations, where an iteration is one presentation of one pattern. In learning to produce the sequences, the network also assigns some value to the don't care unit. This unit is expected to simply maintain the value of the previous pattern; the goal of these simulations is to determine under what conditions the *don't care* unit reverts to the value of the first pattern instead. After 2000 iterations, the training was stopped and the actual output was examined to determine the value taken on by the don't care unit.

c. Results

Results from the first set of simulations show that although the default case is for a don't care unit to maintain the value of the previous output, this previous output is less likely to influence assimilation if it forms a part of a pattern that is strongly dissimilar to the target pattern. In addition, if the pattern two time steps back is strongly similar to the target, the don't care unit will take a value influenced by the corresponding unit in that pattern instead.

These results are given in Graph I, which should be read as follows. Distance along the x-axis measures the similarity between patterns A and C - that is, between the first and third line of each sequence. Column headings refer to sequence in this

145

set, where (i) is the sequence in which pattern A is identical to pattern C (ie, example 8), (ii) is the sequence in which A and C differ on one unit (example 9), up to (vii), in which A and C are completely dissimilar.

```
1.0
(value
 of B)



0.8




0.6




0.-


0.2



0.0
```

i   ii   iii   iv   v   vi   vii
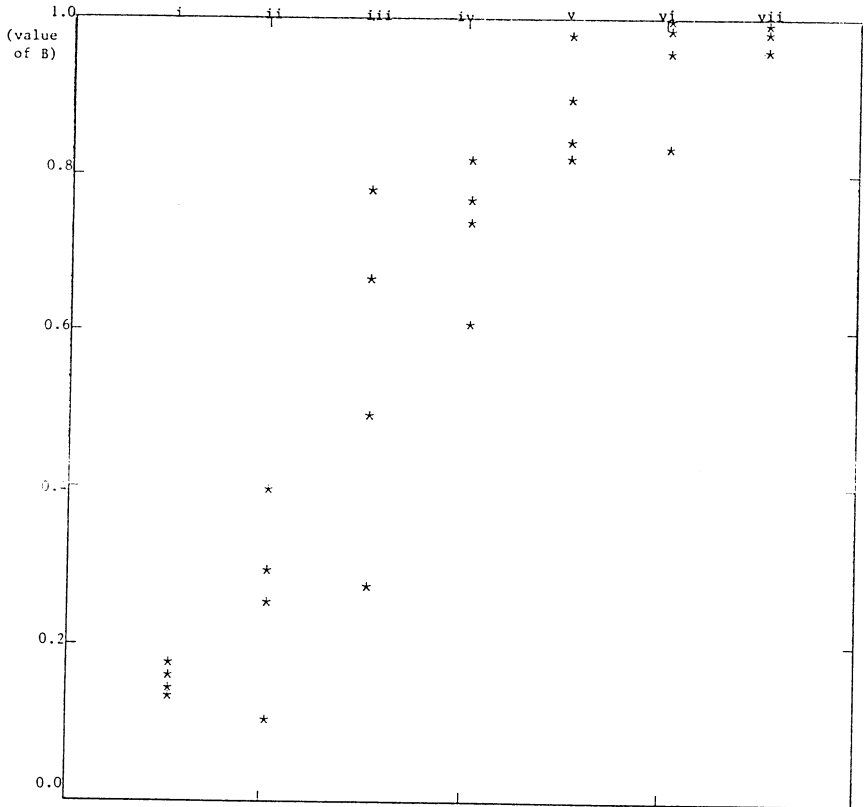
GRAPH I

The row headings on the y-axis give the activation level taken on by the don't care unit after 2000 iterations of learning. On this axis the value for the corresponding unit in B is 1, while the value of that unit in A is 0. Thus the don't care unit will take on a value close to 1 when it is most influenced by B, the immediately preceding output. When the output two time-steps back exerts the most influence, the don't care unit will take on a value close to 0.

In the graph, note in the final three columns that over a number of trials, the *don't care* unit takes on a value of .8 or higher. This is consistent with the interpretation that in these simulations, pattern C has assimilated to B, its immediate temporal predecessor, as might be expected. These are also the sequences in which A is strongly dissimilar to pattern C.

In contrast, the values in (i) and (ii) are much closer to 0, and none is above 0.4. This suggests that when A is markedly similar to C, it is A that exerts the most influence on the assimilation. In the intermediate cases, (iii) and (iv), A is no longer identical or nearly identical to C, yet it is still more similar than is B, the intervening pattern. The conflict between proximity and weakened pattern similarity results in variability in the output, with the don't care unit taking on a range of values depending on initial random conditions.

These results are typical of a pattern which emerged over a number of pattern sets. Graph II gives the results from a second set of sequences, in which B was held constant at a hamming distance of 5 units from C, and A was varied as before. The only difference between the simulations reported above and this set is that B, the second pattern in the sequence, is here slightly more similar to C, the target.

```
1.0          i        ii       iii       iv        v        vi       vii      viii
(value
of B)                                                                   *
                                                     *        *        *
                                              *       *       *
                                                     *
                                                              *        *
0.8

                                        *       *
                                                *       *
                                     *      *
0.6                                         *

0.4                                  *
                               *
                                        *
0.2             *                    *           *
             *
             *
             *           *
                         *
0.0
```
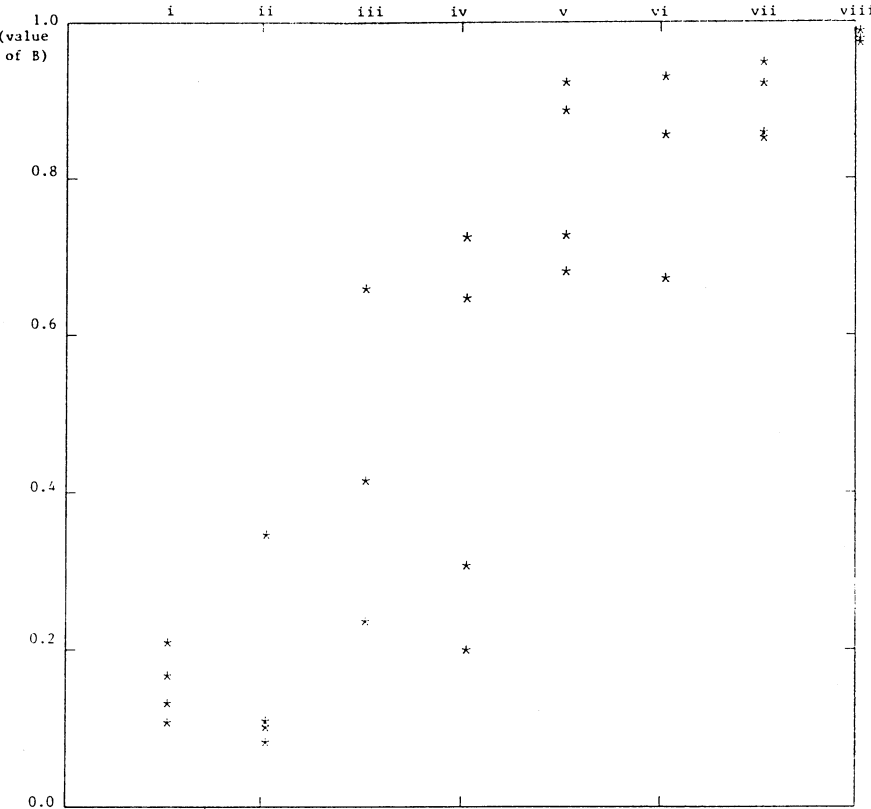
GRAPH II

In this set, the same pattern of results emerges. When A and C are identical, or nearly so, the don't care unit in C takes on a value at the low end of the scale (i, ii). This is the value of the corresponding unit in pattern A. As A becomes increasingly more distinct from C, the resulting values on the don't care unit vary apparently randomly. When A and B are equally similar to C, or if B is more similar to C than is A, the value on the don't care unit is at the high end of the scale, reflecting the influence of pattern B.

Notice also in this graph that there is an eighth column, where all values are clustered near 1, indicating that B was the most influential. This is the output from what will be referred to as the identity condition, where A and B are not only equal in hamming distance from C, but are identical. This result shows that when the two potential trigger patterns are identical or nearly so, the target pattern assimilates to the second of the two in all cases. This is not surprising, given the model. A basic property of these networks is that similar inputs produce similar outputs. Since temporal context is treated here as part of the input, patterns learned in very similar temporal contexts are expected to exhibit very similar behavior.

These simulations were repeated under a number of conditions, with the hamming distance between B and C progressively decreased. The same pattern of results continued to appear, although in an increasingly attenuated form. Consistently, the influence of the second pattern of a sequence is strongest when the first and the third are least similar.

V. Hungarian vowel harmony (II)

This pattern of results shows that in a processor of this sort the similarity structure of output strings across time influences assimilatory behavior. Returning to the Hungarian data, let us consider how the facts of the transparent vowels of that language agree with the behavior of the sequential network.

Here I modeled the behavior of a series of Hungarian words in the same assimilation task. In this case the output sequences were not arbitrary bit strings chosen only for their similarity structure, but vectors corresponding to distinctive feature representations of phonemes. The features used to represent the vowels were *front, low, high,* and *round.* In order to be consistent with the earlier simulations, these were expanded to seven-bit patterns by repeating the last three bits of each string.

(11) Vowel Code:

i	1 0 1 0 0 1 0
e:	1 0 0 0 0 0 0
e	1 1 0 0 1 0 0
ü	1 0 1 1 0 1 1
ö	1 0 0 1 0 0 1
u	0 0 1 1 0 1 1
o	0 0 0 1 0 0 1
a	0 1 0 0 1 0 0

Words being modeled were represented only by their vowels. Each sequence consisted of vectors representing two or more root vowels specified for all features, and a third which represented the vowel of the dative suffix. This was given as a low unrounded vowel unspecified for [front]. As before, lack of specification equated with a don't care condition on the relevant unit.

(12) iker - nek

<pre>
        i   1 0 1 0 0 1 0
        e: 1 0 0 0 1 0 0
        e   * 1 0 0 1 0 0
</pre>

Each pattern was learned separately, as before, for 2000 iterations. At this point the underspecified vowel had taken on a value for [front] influenced either by its immediate predecessor, or by an earlier member of the sequence.

To summarize the results of the earlier simulations, a don't care unit will generally maintain the value on the corresponding unit in the previous output. However, if the immediate predecessor is very dissimilar to the target, it is less likely to trigger assimilation. If the antepenultimate member of the sequence shows a strong similarity to the target, it instead will be chosen as the trigger, and the penultimate member will be ignored. Furthermore, the similarity between the two potential triggers plays a role. If these two are identical, or markedly similar, the target will assimilate to the second of the two in all cases. The current simulations look at how these results explain the real language data.

In the data presented below, the teacher output is given first, followed by the actual output of the network after 2000 iterations of training. The unspecified unit in the teacher is represented with an asterisk (*), and the corresponding output is given in boldface.

For the harmonic roots, both potential triggers have the same value for [front]. This value is straightforwardly maintained onto the don't care unit as a result of the smoothness constraint.

(13) iker - nek

<pre>
        i   1 0 1 0 0 1 0
        e: 1 0 0 0 1 0 0
        e   * 1 0 0 1 0 0
</pre>

0.950	0.047	0.880	0.037	0.118	0.869	0.044
0.966	0.239	0.113	0.034	0.880	0.129	0.044
0.959	0.767	0.062	0.029	0.958	0.043	0.020

(14) kapu - nak

<pre>
a   0 1 0 0 1 0 0
u   0 0 1 1 0 1 1
a   * 1 0 0 1 0 0
</pre>

0.020	0.888	0.109	0.102	0.870	0.104	0.102
0.048	0.197	0.813	0.789	0.220	0.807	0.808
0.038	0.908	0.076	0.133	0.897	0.105	0.109

In the mixed roots, the examples all contain a high front vowel in one syllable and a back vowel elsewhere. In these examples, the vectors representing both the suffix vowel and the back root vowel differ significantly from the root vowel [i]. Thus it is expected that even when [i] immediately precedes the underspecified vowel, it will exert little assimilatory influence. In addition, the first pattern in the string is strongly similar to the target and so is expected to have a strong influence. This is in fact the case, both in the simulation and in the real language data.

(15) taxi - nak

<pre>
a   0 1 0 0 1 0 0
i   1 0 1 0 0 1 0
a   * 1 0 0 1 0 0
</pre>

0.094	0.930	0.074	0.034	0.922	0.066	0.025
0.865	0.130	0.864	0.034	0.144	0.877	0.052
0.115	0.920	0.076	0.028	0.916	0.073	0.020

(16) izom - nak

<pre>
i   1 0 1 0 0 1 0
o   0 0 0 1 0 0 1
a   * 1 0 0 1 0 0
</pre>

0.917	0.084	0.894	0.064	0.087	0.891	0.073
0.098	0.122	0.076	0.840	0.109	0.061	0.855
0.172	0.828	0.090	0.165	0.839	0.100	0.140

However, the situation changes when the root contains a sequence of non-low front vowels. In these examples the target is preceded by a sequence of identical vectors. Here the identity of temporal context is the strongest factor, and the don't care unit is expected to assume the value of its immediate predecessor. Again, this behavior parallels the Hungarian facts.

(17) bronkitis - nek

o	0	0	0	1	0	0	1
i	1	0	1	0	0	1	0
i	1	0	1	0	0	1	0
e	*	1	0	0	1	0	0

0.149	0.098	0.126	0.852	0.098	0.125	0.845
0.888	0.045	0.888	0.098	0.064	0.890	0.112
0.918	0.190	0.817	0.044	0.206	0.833	0.041
0.808	0.607	0.369	0.076	0.589	0.373	0.061

(18) analizis - nek

a	0	1	0	0	1	0	0
a	0	1	0	0	1	0	0
i	1	0	1	0	0	1	0
i	1	0	1	0	0	1	0
e	*	1	0	0	1	0	0

0.036	0.993	0.007	0.011	0.989	0.011	0.040
0.242	0.762	0.242	0.028	0.742	0.237	0.057
0.827	0.174	0.820	0.046	0.185	0.819	0.035
0.897	0.282	0.708	0.033	0.301	0.691	0.016
0.901	0.422	0.573	0.026	0.432	0.569	0.011

VI. Conclusion

To summarize, although the expected pattern in Hungarian is that all vowels of a word will agree in backness, certain front vowels in some environments respect this pattern, and in other environments do not. This more complex behavior is a function of both segmental identity and temporal context. Here I have suggested a processing treatment of the Hungarian facts which predicts harmonic behavior for these vowels on the basis of the overall similarity relationships among the vowels of the word.

A number of factors argue in favor of such an analysis. First, if offers a simpler and more explanatory account of Hungarian data. The vowels which exhibit transparent behavior, and the environments in which this behavior will change, must be stipulated arbitrarily under traditional accounts, while both follow automatically from the account proposed here. Second, this account makes strong claims about the existence of possible harmony patterns. As was demonstrated above, it is a general property of the sequential network that the assimilation process is sensitive to similarity in the temporal context. This predicts substantive constraints on what patterns of harmony may or may not exist. Whether these predictions can be maintained as a general principle requires further research, but available data suggest that they are correct.

Finally, although this account suggests that modifications of the autosegmental treatment of harmony are necessary, it is heavily influenced by the autosegmental notion of assimilation as the spread of a value for a particular phonetic feature. As the simulations demonstrate, a properly constrained treatment of temporal spread correctly predicts those harmonic patterns which exist in Hungarian, while failing to produce non-attested patterns. This result argues strongly in favor of the use of processing models as sources of constraint and explanation which can potentially enrich linguistic theory.

Notes

[1] I am grateful to Jeff Elman, Rob Kluender, Steve Poteet, David Corina, Sanford Schane, Gary Cottrell, Ann Thyme, Errapel Mejiás-Bikandi and Kathleen Carey for useful comments and discussion.

[2] In the literature it is often stated that roots ending in a sequence of high front vowels vacillate in their choice of suffix vowel. However Kontra and Ringen (1987) offer experimental evidence that an overwhelming majority of speakers accepts only front vowels with these stems, and in what follows I assume this to be an accurate statement of the facts.

[3] Results reported here are from a recurrent network with two input (plan) units, seven output and consequently seven state units, and six hidden units. The learning rate in simulations was 0.1; *mu* (the multiplier on the recurrent connections from each state unit to itself) was 0.6, and momentum was set to 0.

Bibliography

Archangeli, Diane. 1984. Underspecification in Yawelmani Phonology and Morphology. MIT PhD dissertation. (Publication forthcoming, New York: Garland Press.)

Archangeli, Diane, and Douglas Pulleyblank. 1986. The Content and Structure of Phonological Representations. ms.

Clements, George N. 1980. Vowel Harmony in Non-linear Generative Phonology. Bloomington, Ind.: IULC.

Goldsmith, John. 1976. Autosegmental Phonology. MIT PhD dissertation. Published 1979, New York: Garland Press.

Goldsmith, John. 1985. Vowel Harmony in Khalkha Mongolian, Yaka, Finnish, and Hungarian. Phonology Yearbook 2:253-273.

Goldsmith, John. 1989. Autosegmental and Metrical Phonology. Oxford: Basil Blackwell.

Hulst, H. van der. 1985. Vowel Harmony in Hungarian: A Comparison of Segmental and Autosegmental Analyses. In van der Hulst and Smith 1985. 267-304.

Hulst, H. van der, and Norval Smith (eds). 1985. Advances in Nonlinear Phonology. Dordrecht: Foris.

Jensen, J.T. 1978. A Reply to "Theoretical Implications of Hungarian Vowel Harmony". Linguistic Inquiry 9:89-97.

Jordan, Michael. 1986. Serial Order: A Parallel Distributed Processing Approach. Institute for Cognitive Science Report #8604. LaJolla: UC San Diego.

Kontra, M., and Catherine Ringen. 1987. Stress and Harmony in Hungarian Loanwords. In Redei 1987. 81-96.

McClelland, James L., David E. Rumelhart, and the PDP Research Group. 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models. Cambridge, Mass.: MIT Press/Bradford Books.

Pulleyblank, Douglas. 1988. Underspecification, the Feature Hierarchy, and Tiv Vowels. Phonology 5:299-326.

Redei, K. (ed). 1987. Studien zur Phonologie und Morphonologie der uralischen Sprachen. Studia Uralica, Wien: Verbund der wissenschaftlichen Gesellschaften Österreichs.

Ringen, Catherine. 1988. Transparency in Hungarian Vowel Harmony. Phonology Yearbook 5:327-42.

Rumelhart, David E., James L. McClelland, and the PDP Research Group. 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations. Cambridge, Mass.: MIT Press/Bradford Books.

Vago, Robert M. 1976. Theoretical Implications of Hungarian Vowel Harmony. Linguistic Inquiry 7:243-263.

nǐ X wǒ Y: A Formal Idiom in Chinese

Zili He
University of Kansas

Linguists have long been intrigued by the principles which govern the distribution and interpretation of proforms in languages. It is traditionally held that the function of personal pronouns is reference--deixis or anaphora (Lyons 1977). This paper investigates some peculiar uses of Chinese first and second person singular pronouns, *wǒ* and *nǐ*, which pose problems for general theories of reference.

Let us begin with a few examples:

(1) *Háizǐmén   nǐ   duǒ   wǒ   shǎn,   bùgǎn   jiànrén.*
    kids   you   dodge I   dodge   not dare see people
    'Every kid was hiding, afraid to be seen.'

(2) *Liǎnggèzǔ   nǐ   cháng wǒ   duǎn,   zěngènòng?*
    two groups you   long   I   short   how to handle
    'The two groups are unequal. How to handle?'

(3) *Fūqī   nǐ   téng   wǒ   aí..*
    husband wife you   love   I   love
    'The couple loves each other...'

(4) *Yīwǎnjiǔ   sìgèrén   nǐ   yī   kǒu   wǒ yī   kǒu -dè*
    one bowl wine four guys you   one   sip   I one   sip -ly
        hē.
        drink
    'Four guys took turns sipping a bowl of wine.'

(5) *Dàjiā   nǐ   pèng   yī   bēi   wǒ   jìng   yī   zhǎn -dè*
    all   you   clink one   glass I   offer one   cup -ly
        guàntā.
        make him drink
    'Everybody was trying to get him drunk, clinking glasses with him one after another.'

(6) *Tāmén   nǐ   kàn   wǒ,   wǒ kàn   nǐ,   méiyǒu   shuōhuà*
    they   you   look at me   I look at you   not   speak
    'They looked at one another, didn't say a word.'

(7) *Yǒudè   jūnzhǎng   zhèngwěi   zài   nǐ   tuī   wǒ*
    some   commanders commissars under you   push   I
        sǎng   -dè   qíngkuàngxià zhànlè   hǎojǐgè   xiǎoshí.
        shove -like circumstances stood   several   hours
    'Some commanders and commissars stood for hours under the circumstances of pushing and shoving (i.e. being pushed and shoved).'

Evidently, the first and second person singular pronouns in 1-7 are NOT used with their ordinary deictic function: *wǒ* is not used by the speaker to refer to himself; *nǐ* is not used by the speaker to refer to his designated hearer.

Next, consider their usage in 8-9 below:

(8)   *Nǐ*   *wú*   *wǒ*   *yǒu,*   *nǐ*   *yǒu*   *wǒ*   *duō ...*
      you   lack   I   have   you   have   I   many
      [Speaking of marketing strategies] 'If one's competitor does not have a
      certain product, one should have it; if one's competitor has it, one should
      have a larger quantity...'

(9)   *Bùshǎo*   *jǔbànjě*   *wufēishì*   *nǐ*   *tāo*   *qián,*   *wǒ pǐng*
      many   organizers   simply are   you   pay   money   I   award
                yǒu;   *nǐ*   *huò*   *mǐng,*   *wǒ*   *dé*   *lì*
                price   you   obtain   fame   I   gain   profit
      [Speaking of commodity contests] 'For many (contest) organizers, it is as
      simple as this: whoever pays (to enter the contests) they will award a prize;
      thus the contestants obtain their fame, and the organizers gain their profit.'

Here, as one can distinctly tell, *wǒ* stands for the one in the interest of which the marketing strategies apply (in 8), and the contest organizers (in 9); and *nǐ*, on the other hand, stands for the competitor (in 8), and the contestants (in 9). In more general terms, although not genuinely deictic, *wǒ* is intended by the speaker to refer to the party taken to be at the center of concern--the 'ego', and *nǐ*, the party taken to be the opposite side--the 'nonego'. I would call this QUASI-DEIXIS: the speaker, instead of being at the deictic center himself, intends that some specific entity be taken as the 'ego', which he may (as in 8) or may not (as in 9) implicitly sides with. Are the uses of *nǐ* and *wǒ* in 1-7 quasi-deictic? I would contend that they are NOT. In 2, for example, one cannot decide which of the two groups is at the center of concern, where the use of *wǒ* is based. Likewise, in 3, there is no justification for the husband, or the wife, being taken as the 'ego' and thus referred to by means of *wǒ* by the speaker. I will not discuss quasi-deixis any further, as it does not fall within the scope of inquiry of this paper.

We have so far assumed that *nǐ* and *wǒ* in 1-7 function somehow referentially. Since they are nondeictic, are they some kind of anaphoric devices, then? At first sight, they appear to be. Although neither makes discrete and specific reference to a particular entity, together *nǐ* and *wǒ* seem to be in coreferential relationship with some appropriate antecedent(s), e.g. in 1 with *háizǐmén* 'kids', in 2 with *liǎnggèzǔ* 'two groups', and in 3 with *fūqī* 'husband and wife'. It will be helpful (though not the task of this paper) to conduct a formal study of the syntactic and discourse constraints for such special anaphor-antecedent linkage. However, a general treatment of the issue in anaphoric terms is not without difficulty. A minor problem is that the anaphoric reference in question is not always determinable on purely structural grounds. Consider 7, which is excerpted from a Chinese government's report on the happenings in Beijing in early June, 1989. (The credibility of the report does not concern us here.) This statement occurs in a discourse context that says something about the siege of a military unit, the attacks on the soldiers, and the great restraint of the army. Hence 7 cannot be taken to mean that the pushing-and-shoving was done by, or between, the commanders and the commissars, although *jūnzhǎng* and *zhèngwěi* are structurally fitting antecedents for *nǐ* and *wǒ*, and are the themes of the passage as well (cf. Sanford 1985). It is only reasonable to conceive that the 'antecedent' of *nǐ* and *wǒ* in this case must be inferentially, rather than grammatically, assigned (cf. Bresnan 1971): the pushing-and-shoving was caused by 'the mob', which is not overtly mentioned.

The major challenge to an anaphoric account comes from the fact that, as a rule, ENTIRE *nǐ X wǒ Y* expressions function UNITARILY as verbals (1, 3, 6), adverbials (4, 5), or adjectivals (2, 7) syntactically, making predications, semantically, of actions, manners, states, etc.; the personal pronouns within these expressions do not play any syntactic role as nominals, and thus do not play any semantic role as referring expressions. Put differently, in a particular utterance (such as 1-7), *nǐ X wǒ Y* has a holistic predicative

function, but neither *nǐ* nor *wǒ* has any separate referential function. Hence, the peculiar uses of the first and second person singular pronouns under discussion cannot be satisfactorily handled in terms of reference (deictic or anaphoric).

Based on 1-7, the following observations can be made:

(i) We have, essentially, a special construction consisting of the parallel structures: *nǐ X wǒ Y*. *X* and *Y* here stand for symmetrical units of verbs (1, 3, 7), adjectives (2), complex nominals (4), or complex verbals (5, 6). When they are complex verbals, each can contain an appropriate element of *wǒ* or *nǐ* (which is the counterpart of the *nǐ* or *wǒ* preceding the complex verbal) to indicate mutuality (see 6).

(ii) Semantically, *X* and *Y* function as equals, in the sense that their interchange does not affect the interpretation of the whole expression. In 1, *duǒ* and *shǎn* are close synonyms, meaning 'to dodge'. Changing from *nǐ duǒ wǒ shǎn* to *nǐ shǎn wǒ duǒ* has no effect on the overall interpretation, namely, 'everyone was hiding'. In other words, the matching of *X* with *nǐ* and of *Y* with *wǒ*, or the other way around, is not substantive, as far as the content is concerned; choices can be made in consideration of conventionality, novelty, tonal patterning, antithesis (relative to some adjacent corresponding expressions), etc. *X* and *Y* can even be antonyms, whose interchange, nonetheless, is not going to alter the reading of the expression as a whole. For instance, in 2 we have an antonymous pair *cháng* 'long' versus *duǎn* 'short'. Either *nǐ cháng wǒ duǎn* or *nǐ duǎn wǒ cháng* carries the same message: 'one long the other short--the two being unequal'.

(iii) Neither *nǐ* nor *wǒ* is intended, in this special construction, to be used referentially. They serve, rather, as a pair of contrastive terms, which suggest certain noncompositional meanings (see iv below). Like *X* and *Y*, *nǐ* and *wǒ* can exchange their positions (again, a matter of choice), with no consequence on the interpretation of the whole expression. Ex. 10 provides an illustration of *nǐ X wǒ Y* and *wǒ X nǐ Y* being juxtaposed antithetically:

(10)   *Yǒuhǎochù, dàjiā   nǐ   zhēng   wǒ   qiǎng;   yǒukùnnán,*
          benefits   all   you   contend   I   contend   difficulties
          dàjiā   wǒ tuī   nǐ   ràng.
          all   I   yield   you   yield
          'For benefits, one vies with another; when it comes to difficulties, one yields to another.'

(iv) The interpretation of a *nǐ X wǒ Y* expression features a crucial noncompositional component, which is intrinsic in the construction and is suggested by the contrastive pair *nǐ* and *wǒ*. Four major subtypes of such noncompositional meaning can be distinguished: (a) each ... the same as the other (1, 7), (b) each ... differently from the other (2), (c) each ... to the other-- mutually (3, 6, 10), and (d) each ... after the other--alternately (4, 5).

In the literature of Chinese linguistics, only sporadic mentions have been made concerning such special usage of personal pronouns (see e.g. Wang 1959, Lü 1985, Zhang 1987). It has been noticed that another pair of basic deictic terms, the demonstrative pronouns *zhè* 'this' and *nà* 'that', have similar uses, as shown in 11:

(11)   *Nèijǐgèrén   nǐ   yī   jù,   wǒ   yī   jù,   zhègè   yī   quán,*
          those people   you   one   word   I   one   word   this   one   fist
          nàgè   yī   jiǎo -dè   zhémó   tā.
          that   one   foot -ly   torment   him
          'Those fellows tormented him, each railing, beating, and kicking him.'

Here, with *nǐ X wǒ Y* and *Zhègè X nàgè Y* in juxtaposition, we can clearly see their similarities in structure, meaning, and function. Interestingly enough, a few other

contrastive deictic pairs are found to serve in some like constructions, as exemplified below:

(12) cǐ   qǐ   bǐ   fú
     this rise that fall
     'rise one after another'
     [cǐ and bǐ : classical forms of 'this' and 'that']

(13) jīntīan pī   Zhāngsān, míngtīan dōu   Lǐsì
     today   criticize Zhangsan   tomorrow   criticize Lisi
     'every day some one, or another, being criticized'

(14) zuǒ sī   yòu   xiǎng
     left think right think
     'ponder from all angles'

(15) lái   zōng qù   jī
     come trace go   trace
     'traces of somebody's whereabouts'

Hence, there might be a _ X _ Y construction where the two slots could be filled in with a pair of contrastive deictic terms, including (in particular) nǐ-wǒ and zhè-nà.

To achieve a more significant generalization, let us shift our attention to the matter of internal structures of Chinese traditional four-syllable idioms (cf. Ma 1983, Xu 1980). We will focus on one of the most prevalent and productive patterns, in which the four elements are in crisscross parallelism structurally and semantically. The following subtypes can be differentiated:

(i)   The two parts (AB - CD) are semantically repetitive.
      (a)   A and C are synonyms.
            (16)   gǎi   tóu   huàn   miàn
                   change head   change   face
                   'change the appearance but not the essence'
      (b)   B and D are synonyms.
            (17)   tóng   qiáng tiě   bì
                   bronze wall   iron   wall
                   'impregnable fortress [figuratively]'
      (c)   A and C, B and D are synonyms.
            (18)   páng   qiāo   cè   jī
                   side   beat   side beat
                   'make oblique references'
(ii) The two parts (AB - CD) are semantically adversative.
      (d)   A and C are antonyms.
            (19)   tóng   chuáng yì   mèng
                   same   bed   different dreams
                   'have different purposes behind the semblance of accord'
      (e)   B and D are antonyms.
            (20)   yǎn   gāo   shǒu   dī
                   eye   high   hand   low
                   'have grandiose aims but puny abilities'
      (f)   A and C, B and D are antonyms.
            (21)   yǒu   tóu   wú   wěi
                   have   head   lack   tail
                   'start something but fail to carry it through'

(iii) The two parts (AB - CD) are neither simply repetitive nor plainly
       adversative.
       (g)   A and C are synonyms; B and D are antonyms.
             (22)   *tiāo   féi   jiǎn   shòu*
                    choose fat   choose   lean
                    'choose whatever is to one's personal advantage'
       (h)   A and C are antonyms; B and D are synonyms.
             (23)   *dōng   lā   xī   chě*
                    east   pull   west   pull
                    'drag in all sorts of irrelevant matters'

As Xu (1980) points out, such pattern-interpretation pairings constitute an important
source of idiomatic productivity. We are most interested in the last group (iii), where each
pattern consists of a pair of synonyms and a pair of antonyms woven in a crisscross
fashion. The synonyms contribute rather directly to the interpretation of the whole;
whereas the antonyms serve as the two poles of a spectrum, suggesting the inclusion of
everything or all possibilities within the range. Ex. 24 is given by Xu as a new idiom
formed on the basis of an old idiomatic pattern (iii.h):

       (24)   *nǐ   zhuī   wǒ gǎn*
              you   catch up   I   catch up
              'overtake each other (in friendly emulation)'

To us, what is significant is not its idiomatic status--being conventionalized and
noncompositional--but the fact that its *nǐ X wǒ Y* form is related to a more general
idiomatic pattern. I would suggest that the nature of the *nǐ X wǒ Y* construction under
discussion could profitably be accounted for in terms of the properties of a more general
idiomatic construction--*A X Z Y*. X and Y are typically near synonyms, which contribute
rather directly to the interpretation of the whole; A and Z are typically a pair of contrastive
terms, (which are associated with the noncompositional component of the interpretation,) in
the general capacity of the two poles of a spectrum, suggesting the inclusion of everything
or all possibilities within the range. Thus, *nǐ X wǒ Y* and similar constructions involving
other deictic pairs could be seen as variants of *A X Z Y*.

In this connection, I would tentatively distinguish two broad categories of *A-Z* terms.
In one class, *A* and *Z* are substantives such as nouns, pronouns, demonstratives, and
localizers (Chao 1968). The associated idiomatic meanings are 'EACH (doing the same
thing, doing a different thing, doing something to the other, or doing something after the
other)', 'EVERYthing', 'EVERYwhere', etc. We have already seen pronouns (*nǐ-wǒ*) and
demonstratives (*zhè-nè*) being used as *A-Z* terms. In 25-29 below, nouns and localizers
serve the same purpose:

       (25)   *gōng   shuō   gōng   yǒulǐ,   pó*
              father-in-law   say   father-in-law   right   mother-in-law
                     shuō   pó   yǒulǐ,
                     say   mother-in-law   right
              'each saying that he himself is right'
       (26)   *shén   bù   zhī   guǐ   bù   jué*
              god   not   know   devil   not   detect
              'without anyone knowing'
       (27)   *tiān   fān   dì   fù*
              heaven   overturn earth   upset
              'everything being turned upside down'

(28) shàng   cuàn   xià   tiào
     up   scurry down   jump
     'run around (on sinister errands)'

(29) lǐ   tiǎo   wài   jué
     inside   poke   outside dig
     'sow dissension everywhere'

In the other class, A and Z are nonsubstantives (Chao 1968), typically adjectives; the associated idiomatic meaning is 'by EVERY means', as in the following examples:

(30) ruǎn   tuō   yìng   kàng
     soft   delay hard   defy
     'employ all sorts of tactics to resist'

(31) jǐn   lán   màn   dǎng
     hasty   block slow   block
     'use various measures to hinder'

(32) sǐ   chán   huó   rào
     dead   twine live   wind
     'keep nagging someone by every means'

(33) hǎo   shuō   dǎi   quàn
     good   talk   bad   persuade
     'try in every possible way to persuade'

(34) fǎn yī quàn zhèng   yī   quàn
     reverse one persuade obverse one persuade
     'persuade from different viewpoints'

(35) héng   dǎ   shùn   suàn
     crosswise   estimate lengthwise calculate
     'plan in view of various factors'

For a start, the focus of the present study is limited to only one A-Z pair, nǐ-wǒ. More extensive and thorough work on various A-Z types is prerequisite to a better understanding of this more general idiomatic construction--A X Z Y.

It is worth mentioning in passing that idiomatic patterns such as A X Z Y or X A Y Z may be prevalent among Sino-Tibetan languages (cf. Ma & Dai 1986). I have come across the following examples:

(36) qʰɔ˦ my˧ tsʰi˥ mi˦
     that sky   this   earth
     'everywhere'
     [Language: Lahu, Branch: Yi]

(37) u˧˩   in˥   kɔ˧˩   in˥
     down   home   up   home
     'neighbors'
     [Language: Achang, Branch: Mian]

(38) ȵtɕʰi˥˩ ta˦ ȵtɕʰi˥˩ sei˥˩
     clean   dead clean   live
     'very clean'
     [Language: Miao, Branch: Miao]

(39) tsʰei˦ vu˧ tsʰei˦ mu˧
     words   go   words   come
     'rumors'
     [Language: Gelao, Branch: ?]

A cross-language survey, inside and outside the Sino-Tibetan Family, of the uses and features of such patterns will greatly enhance our knowledge about the linguistic, cultural, and conceptual aspects of this type of idiomatic constructions.

To understand the nature of *nǐ X wǒ Y*, I have resorted to the theoretical notion of idiomatic construction, or formal idiom (Fillmore, Kay, and O'Connor 1988): a productive construction with special semantic/pragmatic properties not deducible from the general grammar. Moreover, it is important to point out that its idiomaticity, for language users, is realistically a matter of native usage (see the discussions of 'cultural idiomaticity' in He 1989a). This is explicated below from the perspectives of recognition and selection.

The idiomatic character of a *nǐ X wǒ Y* expression such as

(40)   *nǐ*   *qiǎng*   *wǒ*   *duó*
      you   snatch   I   seize
      'vie with each other for something'

is immediately and automatically recognized since it resembles the many popular four-syllable idioms, e.g.

(41)   *lóng*   *zhēng*   *hǔ*   *dòu*
      dragon   contend   tiger   fight
      'a struggle between two evenly-matched opponents'
(42)   *xiān*   *zhǎn*   *hòu*   *zòu*
      first   behead   then   report
      'act first and report afterwards'

What really sounds familiar, and thus gives such expressions an idiomatic ring is their distinctive form--four monosyllabic terms in a two-two rhythm, and in crisscross parallelism structurally and semantically--a form traditionally favored in the culture and characteristic of idiomatic expressions. More significantly, from the psychological point of view, the familiarity of this idiomatic construction could perhaps be attributed to the existence of a number of conventionalized *nǐ X wǒ Y* expressions, which are popularly known and frequently used. For example:

(43)   *nǐ*   *sǐ*   *wǒ*   *huó*
      you   die   I   live
      'life-and-death'
(44)   *ěr*   *yú*   *wǒ*   *zhà*
      you   cheat   I   cheat
      'each cheating the other'
      [*Ěr*: a second person pronoun in ancient Chinese]
(45)   *nǐ*   *yī*   *yán,*   *wǒ*   *yī*   *yǔ*
      you   one   speech   I   one   talk
      'everyone joining in (in the discussion of something)'
(46)   *Nǐ*   *hǎo,*   *wǒ*   *hǎo,*   *dàjiā*   *dōu*   *hǎo.*
      you   good   I   good   everyone   all   good
      'Everyone being happy (at the expense of principles)'

These have probably come to form a paradigm in the memory for the recognition, comprehension, and also free creation of other *nǐ X wǒ Y* expressions.

In speech production, this construction might be selected for its special communicative effects. Stylistically, its use is characteristically informal and colloquial. For example, the

idea of mutuality can be expressed by either *nǐ X wǒ Y* or the adverb *hùxiāng*. Both can be used to talk about common activities in everyday life, with the former felt to be more casual-sounding (compare 47a with 47b below):

(47) *Háizǐmén   zàilóudàolǐ*
      kids   in corridor
    a. *nǐ   tuī   wǒ   sǎng*
        you push   I   shove
    b. *hùxiāng tuī   sǎng*
        mutually push   shove
        'The kids were pushing and shoving each other in the corridor.'

But when it comes to abstract and serious matters in formal speech and writing, only *hùxiāng*. is standard and appropriate, as exemplified by the following set collocations:

(48) *hùxiāng yīcún*
      'be interdependent'
(49) *hùxiāng páichì*
      'be mutually exclusive'
(50) *hùxiāng pèihé*
      'be cooperative'
(51) *hùxiāng chèzhǒu*
      'hold each other back'

Rhetorically, *nǐ X wǒ Y* is capable of producing an undulating rhythm and of evoking vivid imageries. Contrast, for instance, the uses of an adverb *lúnliú* 'alternatively' and *nǐ X wǒ Y* in 52:

(52) *Yīwǎnjiǔ   sìgèrén*
      one bowl wine   four guys
    a. *lúnliúdè   hē*
        alternately drink
    b. *nǐ   yī   kǒu   wǒ yī   kǒu -dè   hē.*
        you one sip   I one sip -ly   drink
        'Four guys took turns sipping a bowl of wine.'

Describing the same event, 52a probably calls forth merely an abstract conception of turn taking, whereas 52b dramatically depicts the scene of the bowl being passed around from mouth to mouth among different individuals.

The polar terms in the *A X Z Y* construction, by virtue of their inherent senses, could produce striking hyperbolic effects, as might be perceived in:

(53) *shén   bù   zhī   guǐ   bù   jué*
      god   not   know devil not detect
      'without anyone knowing'
(54) *yīn   chā   yáng   cuò*
      YIN   mistake YANG mistake
      'a mistake due to a strange combination of all sorts of accidental circumstances'
(55) *tiān   bù   pà   dì   bù   pà*
      heaven not fear earth not fear
      'to fear nothing at all'

There are quite a few familiar *A-Z* pairs, whose usage and interpretation are basically a function of convention, e.g. *nǐ-wǒ* 'you-I', *zuǒ-yòu* 'left-right', *dōng-xī* 'east-west', *héng-shù* 'horizontal-vertical', and *sǐ-huó* 'dead-live'. The construction also encourages individual imagination in the formulation and comprehension of expressions with unusual *A-Z* terms. Such idiomatic creativity is based on the language users' common cultural experience and knowledge, and realized through one's originality in thinking and talent in language use (cf. MacCormac 1986, He 1989b). Here are a few fine examples:

(56) Yījiāzǐ   zhěngtiān   jī   chǎo   é   dòu ...
family   whole day   chicken   quarrel   goose   fight
'(Members of) the family quarreled and fought with one another all the time...'

(57) Háizǐmén   jī   yī zuǐ   yā yī   zuǐ   -dè   dòu   tā
kids   chicken   one mouth   duck one   mouth   -ly   tease   him
'All the kids teased him.'

(58) Tā   wén   chī   xiāngguā,   wǔ   chī   xīguā ...
he   civil   eat   muskmelon   military   eat   watermelon
'He ate muskmelon and watermelon (all day long) ...'

Speaking impressionistically, the imagery of chickens and geese (in 56) highlights incompatibility, of chickens and ducks (in 57), friskiness; the terms *wén-wǔ* civil-military' (in 58) suggest different manners of eating, viz. chewing gently and munching vigorously, respectively. An rhetorical device with remarkable vivifying effect, the *A X Z Y* construction is highly favored by the ordinary folks in daily communication.

The major purpose of this paper is to explore and specify the usage and interpretation of Chinese first and second person singular pronouns in the special construction *nǐ X wǒ Y*. A sample of data mostly collected from written texts has been examined against the general conceptions of deixis, anaphora, reference, and idiomaticity. The main points of the discussions can be summarized as follows:

(i) The uses of personal pronouns under consideration are neither deictic nor quasi-deictic (with a displaced 'ego' arbitrarily assigned by the speaker).

(ii) An anaphoric account in purely structural terms fails in some significant cases where appropriate antecedents, or more precisely, intended referents, have to be inferred from nonlinguistic contexts.

(iii) The most crucial fact is that, as a rule, entire *nǐ X wǒ Y* expressions function unitarily as verbals, adverbials, or adjectivals syntactically--it thus follows that, semantically, whole *nǐ X wǒ Y* expressions make predications of actions, manners, states, etc. The personal pronouns involved are NOT intended to have any separate nominal (syntactic), or referential (semantic), function at all. Hence, the real issue is NOT a matter of reference.

(iv) What we have here is essentially a formal idiom--*nǐ X wǒ Y*: a productive special construction, in which *nǐ* and *wǒ* work as a pair of contrastive terms (with no referential force), while *X* and *Y* are symmetrical structurally and equal semantically.

(v) The interpretation of an *nǐ X wǒ Y* expression includes a characteristic noncompositional component, which is intrinsic in the construction and is associated with the contrastive pair *nǐ* and *wǒ*. Four major subtypes of such idiomatic meaning can be specified: (a) each ... the same as the other, (b) each ... differently from the other, (c) each ... to the other, and (d) each ... after the other.

(vi) I have suggested a more general idiomatic construction *A X Z Y*, under which *nǐ X wǒ Y* is subsumed. This construction may be prevalent among Sino-Tibetan languages.

In short, the uses of *nǐ X wǒ Y* expressions follow not from knowing the general grammatical and pragmatic principles, but from being familiar with the form, meaning, and

effect of a special idiomatic construction--*nǐ X wǒ Y*. Crucially involved in the interpretation process is (instead of the assignment, or inference, of referents for the personal pronouns) the selection of appropriate idiomatic meanings for whole *nǐ X wǒ Y* expressions, guided by the principle of relevance (Sperber & Wilson 1986). I will address this issue in a separate paper.

REFERENCES

Bresnan, Joan. 1971. A note on the notion 'identity of sense anaphora'. Linguistic Inquiry 2.589-97.
Chao, Yuenren. 1968. A grammar of spoken Chinese. Berkeley: University of California Press.
Fillmore, Charles, Paul Kay, and Mary O'Connor. 1988. Regularity and idiomaticity in grammatical constructions. Language 64.501-38.
He, Zili. 1989a. Idiomaticity in language: Cultural awareness and creative usage. Lawrence: University of Kansas dissertation.
---------. 1989b. Creative idiomaticity. Proceedings of the fifteenth annual meeting of the Berkeley Linguistic Society.
Lü, Shuxiang. 1985. Proforms in modern Chinese. Shanghai: Xuelin.
Lyons, John. 1977. Semantics. Vols I & II. Cambridge: Cambridge University Press.
Ma, Guofan. 1983. Idioms. Huerhaote: Inner-Mongolian People's Publisher.
Ma, Xueliang, and Qingsha Dai. 1986. Sino-Tibetan Family. Chinese Encyclopedia: Nationalities, 164-68. Beijing: The Publisher of Chinese Encyclopedia.
MacCormac, Earl. 1986. Metaphors and symbolic activity. Creative Metaphors Journal 1.171-184.
Sanford, Anthony. 1985. Aspects of pronoun interpretation: Evaluation of search formulations of inference. Inferences in text processing, ed. by Gert Rickheit and Hans Strohner, 183-204. Amsterdam: Elsevier Science Publishers B.V.
Sperber, Dan, and Deirdre Wilson. 1986. Relevance: Communication and cognition. Cambridge: Harvard University Press.
Wang, Li. 1959. Modern Chinese grammar. Hong Kong: Chinese Bookstore.
Xu, Zhaoben. 1980. An introduction to Chinese idioms. Beijing: Beijing Publisher.
Zhang, Zhengbiao. 1987. 100 distinctions to make in the study of modern Chinese. Xian: Shanxi People's Publisher.

Information Structure as a Consequence of Word Order Type*

Susan C. Herring
California State University, San Bernardino

1.0 A great deal of confusion persists in the linguistic literature regarding the relationship of word order to the pragmatic notions of 'topic' and 'focus'. It is well-known that word order variation may be pragmatically motivated, or more precisely, that topic and focus-creating operations often involve pre- or post-posing elements within a clause or sentence.[1] Starting with the work on 'functional sentence perspective' by Prague School linguists, and continuing through contemporary generative grammar, a correlation has been noted between 'topic/theme' and sentence-initial position, and 'focus' of a constituent within the 'comment/rheme' by means of rightward movement processes. What is not clear, however, (aside from the obvious problem of what exactly is meant by terms such as 'topic', 'focus', etc.[2]) is whether this correlation necessarily holds true for all languages.

The question is obscured by the fact that a number of researchers readily invoke 'initial topic position' and 'final focus position' as explanations for word order variation in the languages they consider, treating these notions as unexamined primitives. The underlying assumption here -- that the answer is already known -- is a dangerous one, in that it diverts attention away from the kind of empirical cross-linguistic research which needs to be done in order to prove or disprove the hypothesis. Other researchers argue for the universality of the order 'topic-comment' on cognitive and functional grounds. Many of the arguments they present are intuitively appealing; for example, the view that topics come first because they relate thematically to the preceding discourse, while new, focused referents appear later, closer to the subsequent discourse in which they play a part (cf. Hetzron 1975).

Yet anyone who has had experience in working with a variety of language types, especially African or Amerindian languages, is most likely aware that such a claim is false. It is simply not the case that all languages put topics first and focus last; some in fact exhibit the inverse order. This raises the question of what we are to make of the naturalness arguments used to argue in favor of the universality of the topic-focus order. Are some languages pragmatically "less natural" than others? Or should the arguments themselves be reconsidered? Even more interesting, perhaps, is the challenge of accounting for the different pragmatic types -- focus-topic, as opposed to topic-focus order -- attested cross-linguistically. Is there any systematic correlation between pragmatic function and word order, and if so, is it possible to make a general descriptive statement regarding it?

2.0 What I wish to suggest is that the favored positions of topic and focus in any given language are determined by a set of functional and strategic principles, some of which make conflicting predictions. By far the most accurate predictor, as I will demonstrate below, is basic word order type, and specifically, the unmarked relative order of subject and verb (i.e. whether a language is SV or VS). The word order principle is not without exceptions, however. Individual languages as well as entire language types deviate significantly from the predicted patterns in ways which suggest that discourse-functional principles (of the sort proposed in support of the universality of the order topic-focus) occasionally prevail.

In what follows, I identify four different principles which have been hypothesized to influence topic/focus order, and evaluate their predictive accuracy by comparing the predictions they make with the attested orderings of topic and focus constituents in thirty-six languages. The first three principles have all been proposed, either explicitly or implicitly, in the linguistic literature; each, I feel, offers an intuitively plausible functional motivation for pragmatic word ordering.[3] The fourth principle is the basic word order hypothesis mentioned above. The four principles may be stated as follows:

Principle 1: *Given information before new information* (Gundel 1988; cf. also Prague School linguists, especially Firbas 1964 on 'communicative dynamism'). The structure of utterances in discourse builds on what is known, adding new information which moves the communication forward.

Principle 2: *First things first* (Gundel 1988; cf. also the 'newsworthiness' principle, Mithun 1987). Predicts that the most pressing or newsworthy information, including that which is unexpected based on the preceding discourse, will precede less important information in communication.

Principle 3: *Discourse iconicity* (Hetzron 1975 for focus; Prague School linguists for topic). Information ideally is placed as close as possible to the part of the discourse to which it relates; i.e. sentences start with what has already been talked about, and end with what is to be talked about next.

Principle 4: *Word order type* (Creider & Creider 1983; Herring 1989b). Information structure is determined relative to a language's basic word order, as a rhetorical marking strategy. Verb-subject languages tend to order focus (comment) before topic, in contrast with languages of either the SVO or the SOV type.

The first three principles I refer to as 'naturalness' principles, in that they are based on presumed communicative and/or functional naturalness, while the last principle is based on the strategic notion of markedness relative to an accepted norm (in this case, basic word order). The predictions that these principles make regarding the preferred position of topic and focus depend in large part on what exactly is meant by 'topic', and what is meant by 'focus'.

In the present study, I will restrict my observations to two types of topic phenomena -- continuous topics, and shifted topics -- and two focus types -- contrastive focus, and presentational focus.[4] Continuous topics I define as those which carry over from the immediately preceding discourse. They are often pronominalized, or, in some languages, omitted altogether. Shifted topics, on the other hand, are discontinuous, and are often signalled by full lexical NPs and/or by topicalizing expressions such as *as for X, ...* or *X, on the other hand, ...,* etc. As for the two types of focus, contrastive focus highlights a constituent within what is usually considered the 'comment' portion of an utterance, typically in contradistinction to a parallel constituent in a previously-uttered (or implied) clause; an example is an expression introduced by an 'it'-cleft in English, as in *It's John (not Sidney) that I love.* Presentational focus involves the introduction of a referent into a (typically, narrative) discourse; a classic example is an English sentence of the type: *Once upon a time, there was a beautiful princess.* In this paper I leave aside the more general question of what the two topic types and the two focus types have in common; that is, how we might subsume the subtypes into broader definitions of the terms 'topic' and 'focus'.

It is essential to distinguish subtypes of topic and focus in order to determine whether each presents information which is 'given' or 'new'. I understand 'given information' to be that which the speaker may reasonably expect the hearer to be able to access at a particular point in the discourse, based on the situational context and what has been previously said. It is misleading to state (as is often done) that topics in general represent given information; topics are given if they are thematically continuous, but not if they are new or shifted topics. Similarly, the information status of contrastive focus and presentational focus differs, in that arguments presented for the first time (e.g. as participants in a narrative) are completely new, while contrastively focused arguments are already explicitly or implicitly present in the discourse context (an observation which correlates with their tendency to be grammatically definite). The information status of each of these functions is summarized in table 1:

pragmatic function:	information status:
continuous topics	given
shifted topics	new (or reactivated)
contrastive focus	given (or implied)
presentational focus	new

Table 1: Information status of topic and focus sub-types

The four principles listed above make predictions with regard to the position each of the pragmatic functions will favor; that is, whether they will come early or late in the sentence, and whether they will condition pre- or post-posing in marked cases. These predictions are charted in table 2:

	Principle 1	Principle 2	Principle 3	Principle 4 SV/VS
contin. topic	pre-	post-	pre-	pre-/post-
shifted topic	post-	pre-	[pre-]	pre-/post-
contr. focus	pre-	pre-	pre-	post-/pre-
present. focus	post-	[post-]	post-	post-/pre-

Table 2: Predictions made by principles 1-4

Looking down each column vertically, we can see that Principle 1, the given-before-new principle, predicts that given information, such as that in continuous topics and contrastive foci, may be preposed, e.g. in case the targeted element does not already appear in the unmarked topic/focus position in the sentence, or for special emphasis.[5] Shifted topics and presentational foci, which are new or otherwise inaccessible in the local discourse context, should occur later in the sentence, according to this principle.[6] Principle 2, the first-things-first principle, predicts that whatever is most important for the listener to know in order to correctly interpret the sentence should come first; thus contrastive information, such as that expressed in topic shifts and contrastive foci, is likely to be fronted, while predictable, thematic information, such as continuous topics, should occur late. Predictions given in square brackets in the table are those that I feel the principle makes only weakly; thus Principle 2 sanctions the post-posing of presentational foci (in that they are not vital to establishing the interpretation of the rest of the sentence), but does not necessarily restrict them from occurring earlier. Post-posing of presentationally focused arguments is most strongly predicted by Principle 3, the discourse-iconic principle. It is a typical discourse strategy for arguments to be first presented, then commented upon in the following sentence or clause (see Lambrecht 1986); moving them rightward thus places them closer to the context in which they most logically figure. Continuous topics and contrastive foci, on the other hand, are more naturally associated with the previous discourse, and for them pre-posing is predicted. As for shifted topics, they should occur early in case they are contrastive, but the discourse-iconic principle is neutral as to their position if they are new. (In practice, this should depend on whether the new topic is continued in the subsequent utterance or not.) Looking across the table horizontally, we see that the three naturalness principles make conflicting predictions with regard to continuous and shifted topics, but are unanimous in predicting that contrastive focus should appear early in the sentence and presentational focus, late. Thus we might expect that languages, if indeed they are influenced in their pragmatic use of word order by naturalness considerations of this sort, would universally pre-pose contrastive foci and post-pose presented referents, but that there would be some cross-linguistic variability with regard to the positions of the two types of topic.

In contrast with these three principles, the word order principle makes predictions based not on assumed naturalness, but rather on observed tendencies in how certain languages order topic and focus. Creider (1975) was, to my knowledge, the first to note that in verb-initial languages, the focus, or 'rheme', precedes the topic. His observations were based on several VSO African languages of the Nilotic family. Tomlin and Rhodes (1979) made a similar observation for Ojibwa, which has VOS order. In a recent study (Herring 1989b), I compared and contrasted preferred topic and focus positions in 17 languages,[7] 11 of which had VS order, and found that basic word order -- specifically, the order of the subject in relation to the verb -- accurately predicted the order of topic and focus in every case. Generalizing from observations made for a limited set of languages by Creider, Tomlin and Rhodes, and myself, then, the word order principle predicts that topics of either type will be pre-posed in SV languages but post-posed in VS languages, while foci will be post-posed in SV languages but pre-posed in VS languages. The mechanism underlying this principle is presumably a strategic one, based on the goal of achieving salience by the use of word orders which are marked relative to the pragmatically "neutral" order (however this latter is to be defined).

It is evident from table 2 that no two of the four principles make exactly the same set of predictions. Only the word order principle, Principle 4, groups together the two topic categories and the two focus categories. If this principle turns out to represent a general cross-linguistic pattern, it would support the notions of topic and focus as unitary and distinct functional categories in language. Note also that it is only the word order principle which predicts that different languages will order topic and focus differently; the others, based as they are on considerations of functional and communicative naturalness, predict that all languages will prefer a single ordering. (This is true regardless of what one interprets the predictions they make to be.)

3.0 The second part of the analysis consists of comparing the predictions made by the four principles with the attested patterns of topic/focus order in a sampling of typologically diverse languages. I examined data from primary and secondary sources for 36 languages, listed in appendix 1. The languages represent five major word order types: SVO, SOV, VSO, VOS, and OVS;[8] also included are several languages considered to have "free" word order. The procedure I followed in gathering information from secondary sources was to count as evidence of a particular correlation 1) example sentences in which a construction in the language was consistently glossed in English by means of explicit shifted topic expressions such as *As for X, ...* or contrastive focus expressions such as clefts; 2) example sentences for which enough contextual information was provided to determine whether a topic was continuous or shifted, and whether a nominal referent was being presented as a new focus of information (e.g. in narrative texts); and/or 3) explicit remarks on the part of the authors to the effect that a particular word order marks a particular function in the language in question, taking into account differences in terminology, and excluding claims which did not appear to be supported by the actual data presented. There are several obvious problems with this method, not the least of which is that it was not possible to identify all four of the topic/focus subtypes for every language, nor to determine the relative frequency of any construction's use.[9] Even given these inadequacies, however, a number of very distinct patterns emerge, which strongly suggest some sort of overall typological consistency. I will summarize these patterns briefly.

Ignoring, for the moment, individual languages which display exceptions (indicated by asterisks in appendix 1), we may state that topic/focus positions vary predictably according to a language's basic word order type. SVO languages as genetically diverse as English, Mandarin, and Luo all agree in their preference for placing topics of both types in initial -- subject -- position, and focused arguments in post-verbal -- object -- position.[10] SOV languages such as Turkish, Tamil, and Japanese also show a preference for initial topics, but due possibly to a constraint of strict verb-finality in many SOV languages, focus

position is immediately pre-, rather than post-verbal.[11] Thus SOV languages also evidence a correlation between topic and subject, and focus and object. Elements which appear to the <u>right</u> of the finite verb in such languages are backgrounded intonationally and pragmatically, and tend to express highly predictable, thematic information; accordingly, continuous topics may appear sentence-finally. This is also true for SVO languages; consider the English example: *He's an idiot, that guy.* Thus languages of both the SVO and the SOV types generally support the claim that topic precedes focus. We may compare the findings for these two language types with the specific predictions made by our four strategic principles, as given in table 2. This is illustrated in table 3:

Results:		Support predictions made by:			
	SVO/SOV	Principle 1	Principle 2	Principle 3	Principle 4
contin. topic	pre-/pre- (post-/post-)	YES (NO)	NO (YES)	YES (NO)	YES (NO)
shifted topic	pre-/pre-	NO	YES	[YES]	YES
contr. focus	post-/"post-" (pre- /)	NO (YES)	NO (YES)	NO (YES)	YES (NO)
present. foc.	post-/"post-"	YES	[YES]	YES	YES

Table 3: <u>Comparison of predictions with results for SV languages</u>

Looking across table 3, we see that the fronting of continuous topics is correctly predicted by all but principle 2, the first-things-first principle (which does, however, justify post-posed, backgrounded topics of the type mentioned above). Initial position for shifted topics is accurately predicted by all but principle 1, the given-before-new principle. Only principle 4, the word order principle, accurately predicts that contrastive focus may be post-posed (as in the case of WH-cleft constructions), while the post-posing of presented arguments is consistent with the predictions of all four principles. Reading the table vertically, we see that the word order principle most accurately predicts the basic ordering of pragmatic functions of all four types, followed by the discourse-iconic principle, with the given-before-new and the first-things-first principles right in only half of the cases.

What then, of languages in which the order of subject and verb is reversed? A general tendency may be identified for VSO languages, in which focused elements of both types may be moved into pre-verbal position.[12] This is true for example for Jacaltec, Biblical Hebrew, and Nandi. In Podoko, on the other hand, the position of greatest focus is immediately post-verbal, although if anything other than the subject is focused, leftward movement is still required. This situation is in some ways analagous to that of SOV languages with a rigid verb-final constraint, in that the initial or final verb acts as a barrier to focus movement, such that the focused element moves as far as it can in the direction of the verb before being blocked. As for topics, the two subtypes split, with continuous topics appearing post-verbally, and shifted topics pre-verbally, in virtually all of the VSO languages in my sample. VOS languages such as Ojibwa and Malagasy pattern the same as do VSO languages; in both, focused elements and shifted topics may be pre-posed, while continuous topics remain post-verbal. The OVS languages in the sample, Hixkaryana and Makushi, also pattern like verb-initial languages for all four functions, as do the three "free" word order languages discussed in Mithun (1987), which exhibit a statistical preference, albeit slight, for OVS order. Generalizing, we may state that in languages in which the basic order is verb-subject, the order focus-topic is preferred, unless the topic is new or contrastive, in which case it appears initially. These results are matched with the predictions of the four principles in table 4 below:

Results:		Support predictions made by:			
	VSO/VOS/ OVS	Principle 1	Principle 2	Principle 3	Principle 4
contin. topic	post	NO	YES	NO	YES
shifted topic	pre-	NO	YES	[YES]	NO
contr. focus	pre- ("pre-")	YES (YES)	YES (YES)	YES (YES)	YES (YES)
present. foc.	pre-	NO	[NO]	NO	YES

Table 4: <u>Comparison of predictions with results for VS languages</u>

It can be seen from table 4 that the three naturalness principles fare no better or worse for VS languages than for SV languages -- that is, they make correct predictions only about half the time. The word order principle makes one incorrect prediction here, namely that shifted topics, like continuous topics, should be post-posed. Even given this, it is by far the most accurate predictor of the actual orderings of any of the principles considered.

However it is not the case that naturalness principles have nothing to say about the pragmatic functions of word order. Languages of <u>all</u> types are consistent in putting shifted topics first, even against the predictions of the word order principle in VS languages. This ordering is conditioned most significantly by the first-things-first principle, and is probably the one true universal of pragmatic ordering. When the topic of a sentence shifts or is otherwise unpredictable, it must be explicitly stated at the outset, lest the hearer misinterpret what the sentence is about. (Similar observations have been made by Creider & Creider (1983), Gundel (1988), and Mithun (1987).)

Another area in which naturalness principles hold sway is with contrastive focus. Post-posed constrastive focus such as is found in SV languages is counter-indicated by all three naturalness principles. While on the one hand this is strong evidence in favor of the word order hypothesis, it is also not without exception. As an example, consider the problem of it-clefts in English, the counterpart of which is found in numerous SVO languages. By one analysis, it-cleft constructions such as *It's John that I love* <u>pre</u>- pose the object NP, *John*, whereas by another analysis, *It's John* is a type of presentative construction in which the order of subject and verb has been inverted, making *John* the focus of a <u>post</u>-posing process. Regardless of how one resolves this controversy, there is evidence from other SVO languages which is even more problematic. In Kru (Marchese 1983), the focused element in it-type clefts appears in sentence-initial position, followed by a copula functioning as focus marker. Thus there is no possible analysis whereby such focused elements could be viewed as anything other than fronted. Similarly, nominalized clefts (e.g. *the one who was sacrificed was Abraham*) occur more often in "reversed" order (*ABRAHAM was the one who was sacrificed*) -- that is to say, with the focused element first -- than in normal order in Lele, according to Simons (1982). This suggests that there is, in the case of SV languages, a tension between the word order principle and the naturalness principles with regard to contrastive focus, and that it is not always the word order principle which prevails. Other exceptions can be explained similarly, i.e. as the result of a conflict between predictions made by different functional principles.

4.0 If, as these results indicate, word order type is a valid predictor of pragmatic ordering cross-linguistically, the question naturally arises as to what motivates the strategy in the first place. Although in the present paper I have made a heuristic division between the word order principle on the one hand, and commonly invoked naturalness principles on the other, I do not intend to imply by this that the word order principle is in any way <u>unnatural</u>. On the contrary, given its apparently universal distribution, we must suspect that it is driven by basic and easily-accessible strategic principles.

I have suggested here that the notion of pragmatic markedness plays a key role in the analysis of the word order principle. As a strategy which produces marked word orders to express emphasis or contrast, it is iconic, in that it draws attention to itself by virtue of its deviation from the expected norm. This explanation leaves unanswered, however, the question of why the strategy takes as its point of departure the relationship between subject and verb, rather than that between verb and object (as in traditional Greenbergian typological analysis) or between subject and object.

The answer lies, I believe, in the interaction of two asymmetries in discourse patterns. First, there is a well-attested cross-linguistic correlation between grammatical subject and pragmatic topic; subjects are prominent on hierarchies of animacy and agentivity (Silverstein 1976), while topics are prominent in discourse as (typically) animate and agentive thematic participants (Givón 1983). The correlation between direct object and focus is much weaker, however. First, not every sentence has an object, but it may still have a focus, since focus may be a property of predicates and of sentences as a whole, as well as of 'narrow' focus nominal arguments (Lambrecht 1986). From the point of view of discourse, although DuBois (1987) characterizes direct object as a preferred argument role for the introduction of new discourse referents, Herring (1989a) presents evidence to the effect that intransitive subjects are significantly more strongly preferred in this function, particularly when the referent is animate. This further weakens the correlation of direct object with pragmatic focus. On the basis of this asymmetry, then, we may speculate that a language user is more likely to notice a correlation between grammatical argument position and pragmatic function in the case of subject/topic, where a stronger correlation exists. Following this line of reasoning further, the language user then generalizes canonical subject position (whatever it happens to be in his or her language) as topic position, allowing for the possibility of topicalizing arguments other than grammatical subject. As for what determines focus position, it is likely that the relative complementarity of the two pragmatic functions, and the fact that it is sufficient to order them with respect to one another, combine to select a focus position which requires movement (in those cases when movement is required) in the direction opposite of that for topicalization. This tendency is no doubt supported by the weaker, but still perceptible, correlation between 'new' focused referents and direct object position.

The second asymmetry involves the discourse motivations for creating marked word orders. In narrative discourse, thematic participants (i.e. referents which persist or recur as utterance topics across continuous stretches of discourse) are typically animate and agentive, as mentioned above, and function as the grammatical subject of the clause in which they appear. While this pattern is the most statistically frequent, a different situation nevertheless arises when these same referents are introduced into the discourse for the first time. At this moment, they are not yet topics; rather, they are typically accorded some sort of presentational focus. In order to _focus_ a subject NP, however, it is necessary to either mark it as exceptional by means of intonation or explicit focus morphology, or else to move it out of canonical topic position. In such cases, since presentational verbs tend to be intransitive (_be, live, stand, sit, come, go,_ etc.), the most consistently available word order device is one which permutes subject and verb relative to one another, as in _Once upon a time, there was a princess_ -type sentences in English. Thus the marked order of subject and verb addresses a clear functional need. The same is not true of the relation of verb to object. A parallel explanation would require us to identify a strong discourse motivation for topicalizing objects, yet it is difficult to imagine what such a motivation might be. To the contrary, topicalized objects (with the possible exception of objects promoted to subject via passivization) are rare in narrative, no doubt due to the predominance of the tendency described above for topics to be expressed grammatically as subjects. Add to this the observation that referents occurring in object position are frequently inanimate props which lack thematic persistence, and the apparent motivation for topicalizing objects decreases even further. This leaves us with the third logical possibility mentioned above, namely that the word order principle takes as its point of reference the relative order of subject and

object. While many of the word order permutations considered do in effect reverse the order of subject and object, direct objects -- overt or otherwise -- are often unavailable, since actual discourse makes use of intransitive as well as transitive verbs (Herring 1988 notes that 70% of all finite verbs in a Tamil oral narrative corpus are intransitive). Most sentences have a predicate, however. If the relationship of subject to object were the primary determinant of the pragmatic word order strategy described here, it would require speakers to ignore a consistently available relationship (subject/predicate) in favor of one that is attested only part of the time. This seems inherently more difficult to motivate, and in the absence of any supporting evidence, I dismiss this possibility as well. Thus not only is the correlation of subject to topic more salient than the correlation of object to focus, but there is a clear discourse motivation, in narrative at least, for permuting the order of S and V, whereas in the case of the relationship of O to V, or S to O, such a motivation is lacking. Both asymmetries point to the centrality of S: S has a clearly identifiable pragmatic function, and the relationship of S to V -- the two sentence elements which are typically minimally present -- allows for the inference of relative order. While these observations are necessarily speculative and leave many particulars of the process unarticulated, they suggest, I believe, some plausible ways in which the word order principle can be motivated by relatively simple and universal mechanisms.

5.0 I conclude with an appeal to linguists interested in topic/focus phenomena and/or linguistic typology to take up the challenge of verifying -- or disproving -- the analysis presented here. The program, as I see it, will crucially involve careful investigation of discourse-pragmatic patterns in less familiar languages, especially those in which the basic order is VS. The full range of topic and focus subtypes needs to be considered as well, if we are to arrive at an adequate understanding of the nature of these categories, and of their typological correlates. We have nothing to lose by such an effort, since no other existing account is able to explain cross-linguistic pragmatic word order variation. We stand to gain a powerful predictive tool.

ENDNOTES

* I am grateful to John Paolillo, Eric Pederson, and Randy LaPolla for comments and discussion which contributed significantly to the present paper.

[1] Languages also employ other means of indicating topic/focus functions, e.g. intonation and explicit topic/focus morphology, but due to space limitations, these will not be considered further here.

[2] For example, in the literature on Philippine languages, the term 'focus' refers to a grammatical category which is definite, and which more closely resembles the notion of topic in non-Philippine studies, while the term 'topic' is used to describe elements in contrastive focus (cf. Naylor 1975).

[3] Other functional principles have no doubt been proposed, or will be proposed in the future, which might profitably be included in a comparative endeavor of this sort as well. Herring (1989b) mentions several cognitive principles hypothesized to influence topic-focus order; this is clearly an area requiring further study.

[4] There are other types of focus that would need to be included in a broader study, for example predicate focus and sentence focus (cf. Lambrecht 1986), and within the narrow focus range, WH- focus. It is also possible that the set of phenomena I have grouped together under 'contrastive focus' does not really constitute a functionally homogenous category; we might, for instance, wish to separate out WH- type clefts from it-type clefts, if it could be determined that there was a consistent difference in the functions of the two across languages which have both types.

5 It is important to note that in no language would we predict that pre-posing or post-posing is essential to the expression of topic or focus functions. Rather, movement operations constitute a marked option which a language may have available, and which may be strongly favored under certain grammatical-pragmatic conditions. For example, topicalized objects are more likely to be moved than topicalized subjects, since in most languages, the S position is the normal (unmarked) topic position, while the O position is associated with focus. For similar reasons, focused subjects appear to condition movement more often than focused objects.

6 Jeanette Gundel (personal communication) disagrees that the given-before-new principle makes the prediction that shifted (new) topics would be post-posed, on the grounds that the topic, even if shifted, should always be considered 'given' relative to the comment or focus. This view appears to confuse the topic-comment relationship with the given-new status of referents. While it is indeed the case that no language (to my knowledge) post-poses shifted topics -- a fact which I account for in terms of the overriding of the word order principle by the naturalness principles -- I believe that it is, at least, theoretically possible. It is certainly possible for topics to be new, and comments given. I have in mind discourses such as the following:

1) A: *I'm so proud of my David; he just graduated from college magna cum laude.*
   B: *Now my Julie, she graduated from college magna cum laude. And she got a job working for a senator in Washington!*

2) A: *I'm beat. I really need to get some sleep.*
   B: *You need to get some sleep? I need to get some sleep!*

In B's initial utterance in example 1, the comment (*graduated from college magna cum laude*) is given; what is new is the shift in topic (from *my David/he* to *my Julie/she*). Similarly in 2, B's contrastive assertion *I need to get some sleep* relates a given comment to a new topic. (Notice that these constructions cannot be analyzed as contrastive focus constructions; the corresponding it-cleft focus constructions force a narrower interpretation; e.g. *It's my Julie who graduated from college magna cum laude* seems to deny that *David graduated from college magna cum laude*). New topics with given comments appear to be characteristic of a "one-upmanship" discourse style.

7 The 17 languages surveyed in Herring (1989) include Ojibwa, but do not include any of the languages discussed by Creider, in that I was not aware of Creider's work when I undertook my initial study.

8 As of this writing, I have not been able to gather enough information on pragmatic ordering in OSV languages to include them as a category.

9 I excluded from a larger sample those languages for which fewer than 3 of the 4 topic/focus sub-types could be determined from the available sources.

10 It is not uncommon for SVO languages to have two types of contrastive focus constructions, as exemplified by the two types of cleft (WH- and it-) in English. In these languages, one construction post-poses the focused element, while another appears to pre-pose it; hence the 'pre-' in parentheses in table 3.

11 I designate contrastive focus in these languages as "post"-posing, since in the case of all but direct object focus, rightward movement is involved. Interestingly, the strict verb-final constraint disappears in nominalized predicate constructions in SOV languages such as Tamil and Sinhala, where the focused constituent is preferentially post-verbal, as in the Sinhala example below (Gair 1970; 136):

hæmadaama putaa kanne   nikan bat
every-day son   eat-pr-emph plain rice
'It is plain rice (that) the son eats every day.'

This observation requires further verification for presentational focus, however, in that presentational data are rare in my VSO language sources.

Appendix 1: Languages included in sample

(* indicates deviation from word order principle)

SVO
English
Spanish (Bolinger 1952)
French (Lambrecht 1986, 1987)
Mod. Hebrew (Berman 1980; Givón 1976)
Mandarin (Li & Thompson 1981)
Luo (Creider & Creider 1983)
Kimatuumbi (Odden 1984)
Pulaar (Fagerberg 1983)
Lele (Simons 1982)
Kru* (Marchese 1983)

SOV
Tamil
Sinhala (Gair 1970, 1986)
Turkish (Erguvanli 1984)
Japanese (Hinds 1973; Kuno 1972)
Korean (Kim 1985, 1988)
Hungarian (Kiss 1977)
Sherpa (Givón 1984)

"Free" Word Order: topic-focus
Gavião (SOV/SVO) (Stute 1986)
Xavante* (OSV/SOV) (Burgess 1986)

VSO
Jacaltec (Craig 1977)
Biblical Hebrew (Givón 1984)
Welsh (Williams 1980)
Irish (Bammesberger 1982)
Tagalog (Naylor 1975)
Nandi (Creider 1975; Cr. & Creider 1983)
Podoko (Jarvis 1981)
Turkana (Dimmendaal 1983)
Classical Arabic* (Beeston 1970, 1974)

VOS
Ojibwa (Tomlin & Rhodes 1979)
Malagasy (Keenan 1976)

OVS
Hixkaryana (Derbyshire 1977)
Makushi (Derbyshire & Pullum 1981)

"Free" Word Order: focus-topic
Cayuga (OVS) (Mithun 1987)
Ngandi (OVS) (Mithun 1987)
Coos (OVS) (Mithun 1987)
Ute (SOV/OVS) (Givón 1984)

REFERENCES

Bammesberger, Alfred. 1982. **Essentials of Modern Irish**. Heidelberg: Carl Winter.
Beeston, A.F.L. 1970. **The Arabic Language Today**. London: Hutchinson.
Beeston, A.F.L. 1974. "Embedding of the theme-predicate structure in Arabic." **Language** 50.3; 474-477.
Berman, Ruth. 1980. "The case of an (S)VO language." **Language** 56:4; 759-776.
Bolinger, Dwight. 1952. "Linear modification." **Publications of the Modern Language Association** 67; 1117-1144.
Burgess, Eunice. 1986. "Focus and topic in Xavante." In **Sentence Initial Devices**, Joseph Grimes (ed.); 27-41. Dallas: Summer Institute of Linguistics.
Craig, Colette G. 1977. **The Structure of Jacaltec**. Austin: University of Texas Press.
Creider, Chet A. 1975. "Thematization and word order." Paper presented at the annual meeting of the Linguistic Society of America.
Creider, Chet A. and Jane T. Creider. 1983. "Topic-comment relations in a verb-initial language." **Journal of African Languages and Linguistics** 5; 1-15.

Derbyshire, Desmond. 1977. "Word order, universals, and the existence of OVS languages." **Linguistic Inquiry** 8.3; 590-597.

Derbyshire, Desmond & Geoffrey K. Pullum. 1981. "Object-initial languages." **IJAL** 47.3; 192-214.

Dimmendaal, Gerrit J. 1983. "Turkana as a verb-initial language." **Journal of African Languages and Linguistics** 5; 17-44.

DuBois, John W. 1987. "The discourse basis of ergativity." **Language** 63.4; 805-855.

Erguvanli, Eser E. 1984. **The Function of Word Order in Turkish Grammar. University of California Publications in Linguistics, Vol. 106.** Berkeley and Los Angeles: University of California Press.

Fagerberg, Sonja. 1983. "Discourse Strategies in Pulaar: the use of focus." **Studies in African Linguistics** 14.2; 141-157.

Firbas, Jan. 1964. "On defining the theme in functional sentence analysis." **Travaux Linguistiques de Prague** 1; 225-240.

Gair, James W. 1970. **Colloquial Sinhalese Clause Structures. Janua Linguarum Series Practica** 83. The Hague: Mouton.

Gair, James W. 1986. "Sinhala focused sentences: naturalization of a calque." In **Structure, Convergence and Diglossia**, Bh. Krishnamurti (ed.); 147-164. Delhi: Motilal Banarsidass.

Givón, Talmy. 1976. "On the V-S word-order in Israeli Hebrew: pragmatics and typological change." In **Studies in Modern Hebrew Syntax and Semantics**, P. Cole (ed.); 153-181. Amsterdam: North Holland.

Givón, Talmy. 1983. **Topic Continuity in Discourse: A Quantitative Cross-Language Study.** Amsterdam: John Benjamins.

Givón, Talmy. 1984. **Syntax: A Functional-Typological Introduction. Vol.1.** Amsterdam: John Benjamins.

Gundel, Jeanette K. 1988. "Universals of topic-comment structure." In **Studies in Syntactic Typology**, Michael Hammond, Edith Moravcsik and Jessica R. Wirth (eds.); 209-239. Amsterdam: John Benjamins.

Herring, Susan C. 1988. "Aspect as a discourse category in Tamil." **BLS** 14; 280-292.

Herring, Susan C. 1989a. "Verbless presentation and the discourse basis of ergativity." **CLS** 25; 123-137.

Herring, Susan C. 1989b. "Is sentence-final 'focus' position universal?" Paper presented at the annual meeting of the Linguistic Society of America.

Hetzron, Robert. 1975. "The presentative movement, or why the ideal word order is V.S.O.P." In **Word Order and Word Order Change**, Charles N. Li (ed.); 347-388. Austin: University of Texas Press.

Hinds, John. 1973. "Theme and rheme in Japanese." **Seishin Studies** 41; 1-13.

Jarvis, Elizabeth. 1981. "Some considerations in establishing the basic word order of Podoko." **Studies in African Linguistics** 12.2; 155-167.

Keenan, Edward L. 1976. "Remarkable subjects in Malagasy." In **Subject and Topic;** Charles N. Li (ed.); 337-384. New York: Academic.

Kim, Alan Hyun-Oak. 1985. "The function of linear order in Korean syntax." **Harvard Studies in Korean Linguistics;** 154-167. Cambridge: Harvard University.

Kim, Alan Hyun-Oak. 1988. "Preverbal focusing and type XXIII languages." In **Studies in Syntactic Typology**, Michael Hammond, Edith Moravcsik and Jessica R. Wirth (eds.); 147-169. Amsterdam: John Benjamins.

Kiss, H. 1976. "Topic and focus in Hungarian syntax." **Montreal Working Papers in Linguistics** 8; 1-42.

Kuno, Susumo. 1972. "Functional sentence perspective: a case study from Japanese and English." **Linguistic Inquiry** 3; 269-320.

Lambrecht, Knud. 1986. "Topic, focus, and the grammar of Spoken French". Ph.D. dissertation, University of California, Berkeley.

Lambrecht, Knud. 1987. "On the status of SVO sentences in French discourse". In **Coherence and Grounding in Discourse,** Russell S. Tomlin (ed.); 217-261. Amsterdam: John Benjamins.

Li, Charles N. & Sandra A. Thompson. 1981. **Mandarin Chinese: A Functional Reference Grammar.** Berkeley and Los Angeles: University of California Press.

Marchese, Lynell. 1983. "On assertive focus and the inherent focus nature of negatives and imperatives: evidence from Kru." **Journal of African Languages and Linguistics 5;** 115-129.

Mithun, Marianne. 1987. "Is basic word order universal?" In **Coherence and Grounding in Discourse,** Russell S. Tomlin (ed.); 281-299. Amsterdam: John Benjamins.

Naylor, Paz B. 1975. "Topic, focus, and emphasis in the Tagalog verbal clause." **Oceanic Linguistics 14;** 12-79.

Odden, David. 1984. "Formal correlates of focusing in Kimatuumbi." **Studies in African Linguistics 15.3;** 225-299.

Silverstein, Michael. 1976. "Hierarchy of features and ergativity." In **Grammatical Categories in Australian Languages,** Robert Dixon (ed.); 112-171. New York: Humanities Press.

Simons, Pamela. 1982. "Nè...ba marking in Lele: a cleft construction?" **Studies in African Linguistics 13.3;** 217-229.

Stute, Horst. 1986. "Constituent order, cohesion and staging in Gavião." In **Sentence Initial Devices,** Joseph Grimes (ed.); 7-25. Dallas: Summer Institute of Linguistics.

Tomlin, Russell and Richard Rhodes. 1979. "An introduction to information distribution in Ojibwa." **CLS 15;** 307-320.

Williams, Stephen J. 1980. **A Welsh Grammar.** Cardiff: University of Wales Press.

Quantifier Float in Korean*

Ki-Sun Hong
Stanford University

0. Introduction

The purpose of this paper is to investigate the so-called quantifier float construction in Korean (Shibatani 1977, Gerdts 1985, Lee 1989, among others). I will first examine previous analyses, concentrating on Gerdts' analysis in terms of constituent structure. Second, I will present two kinds of problems in her proposal: one conceptual and the other empirical. Third, I will propose an alternative view. I will show that this construction is subject to a more general condition on the redundancy of semantic information, which is independently required in Korean grammar.

1. Previous Analyses

Quantifier phrases in Korean can appear outside of the noun phrases which they modify, as in (1-2).[1]

(1) a. sey   haksayng-i   o-ass-ta.
     three student-nom come-past-indicative
     'Three students came.'
   b. haksayng-i *seys-i*   o-ass-ta.
     students-nom three-nom come-past-indicative
(2) a. nay-ka sey haksayng-ul manna-ass-ta.
     I-nom three student-acc   meet-past-indicative
     'I met three students.'
   b. nay-ka haksayng-ul *seys-ul*   manna-ass-ta.
     I-nom student-acc three-acc meet-past-indicative

They form a separate constituent from those NPs, as evidenced by adverb insertion as in (3).

(3) a. haksayng-i *ecey*   seys-i   o-ass-ta.
     student-nom yesterday three-nom come-past-indicative
     'Three students came yesterday.'
   b. nay-ka haksayng-ul *wuyenhi* seys-ul   manna-ass-ta.
     I-nom student-acc   by chance three-acc meet-past-indicative
     'I happened to meet three students.'

Following some studies on quantifier float (O'Grady 1982, Belletti 1982, Gerdts 1985), I consider these quantifiers as base-generated anaphors, although it is not the main concern of this paper.

This construction is not always possible: when the relevant NP is marked with a case neither nominative nor accusative (e.g., *eykey* 'to', *eykeyse* 'from', *eyse* 'at', etc.),[2] it cannot antecede a quantifier phrase.

(4) a. nay-ka sey haksayng-eykey kong-ul tenci-ess-ta.
     I-nom three student-to   ball-acc throw-past-indicative
     'I threw a ball to three students.'
   b.*nay-ka haksayng-eykey *seys-eykey* kong-ul tenci-ess-ta.
     I-nom student-to   three-to   ball-acc   throw-past-indicative

(5) a. nay-ka sey hakkyo-loputhe phyenci-lul pat-ass-ta.
     I-nom three school-from letter-acc receive-past-indicative
     'I received letters from three schools.'
  b.*nay-ka hakkyo-loputhe *seys-uloputhe* phyenci-lul pat-ass-ta.
     I-nom school-from three-from letter-acc receive-past-indicative

This contrast between (1-2) and (4-5) leads Shibatani (1977) to claim that quantifier float is sensitive to surface case: quantifiers can be anteceded only by NPs marked nominative or accusative.

This analysis implies that the case of the antecedent NP and that of the quantifier should agree. Gerdts (1985) argues against this proposal, mostly based on complex sentences where this agreement view fails to provide a correct prediction. First, in a subject-to-object raising construction, the raised object can antecede the quantifier phrase. Either nominative or accusative case can appear on the quantifier, as in (6).

(6) a. John-i sey haksayng-ul [chencay-la-ko] mit-nunta.
     nom three student-acc genius-be-comp believe-indicative
     'John believes three students to be geniuses.'
  b. John-i haksayng-ul *seys-i/ul* chencay-la-ko mit-nunta.
     nom student-acc three-nom/acc genius-be-comp believe-indicative
                                    (Gerdts (17-18))

Under the surface case analysis, it is predicted that the quantifier phrase is marked only in accusative case: it cannot be marked in nominative since there is no antecedent with this surface case. Second, the same kind of problem is observed in a causative sentence, as in (7). The surface case analysis cannot account for the nominative case marking on the quantifier phrase, since there is no possible antecedent with that case.

(7) a. nay-ka sey haksayng-eykey [ttena-key] ha-yess-ta.
     I-nom three student-to leave-comp make-past-indicative
     'I made three students leave.'
  b. nay-ka haksayng-eykey *seys-i* ttena-key ha-yess-ta.
     I-nom student-to three-nom leave-comp make-past-indicative (Gerdts (19))

Third, a topic sentence provides us with further counterevidence. Here again, there is no antecedent sanctioning the quantifier phrase in accusative case.

(8) Chayk-un nay-ka *seys-ul* sa-ass-ta.
    book-topic I-nom three-acc buy-past-indicative
    'As for books, I bought three.'   (Gerdts (20))

On the basis of these arguments, Gerdts claims that the account in terms of surface case cannot be maintained. She notes that these data are not tenable with the grammatical relations account either, which dictates that only subjects and objects can antecede quantifier phrases.

(9) a. haksayng-i sonyen-tul-ul chayk-ul *seys-ul* cwu-ess-ta.
     student-nom boy-plural-acc book-acc three-acc give-past-indicative
     'The student gave the boys three books.'
  b. ku cikkong-i sonkalak-i kikye-ey *seys-i* calli-ess-ta.
     the worker-nom finger-nom machine-loc three-nom cut-past-indicative
     'Three fingers of the worker were cut in the machine.'   (Gerdts (21-22))

Chayk 'book' in (9a) and *sonkalak* 'finger' in (9b) sanction quantifier phrases, although they are chomeurs within her theory.[3] Accordingly, the grammatical relations account fails to provide an adequate explanation.

Instead of these two accounts, which have turned out to be inadequate, Gerdts proposes an approach based upon constituent structure, as in (10).

(10) Only nominals which are not contained within a Postpositional Phrase can antecede
     Quantifiers.
                                                            (Gerdts (27))

This statement is based on two assumptions. First, in Korean, only nominative and accusative case are inflectional suffixes attached to a nominal in the lexicon, while other cases are postpositions which form a PP with the preceding NP in the syntax.[4] This explains why only NPs in nominative or in accusative case can antecede a quantifier phrase in (1-5) and (9). Second, an empty NP is required to explain (6-8). In those constructions, there should be an empty NP which bears some function in the embedded clause, and this NP antecedes a quantifier phrase according to (10).

2. Problems in Gerdts

In spite of a consistent account of given data, Gerdts' analysis leaves two problems. First, contrary to her claim, NPs marked in non-nominative/accusative case markers as in (4-5) may sanction quantifier float, even in simple sentences. Second, lexical phonology (Kiparsky 1982, 1985, Mohanan 1982) provides us with evidence that all the case markers should be inflectional suffixes attached to a noun in the lexicon (Y.Cho 1988).[5] Accordingly, we cannot rely on differences in morphological status of the two classes of case markers. Let us consider each problem in detail.

First, NPs marked with non-nominative/accusative case markers can antecede quantifier phrases in two cases: (i) when a delimiting particle (*man* 'only', *to* 'also', *cocha* 'even', etc.; for discussion of so-called delimiters, see Yang 1972)[6] is attached to either the noun phrase or the quantifier phrase, as in (11), and (ii) when either the noun phrase or the quantifier phrase gets stress for contrastive reading, as in (12). As far as case marking is concerned, all of the sentences include NPs in dative case. However, delimiters or stress make them perfectly grammatical.

(11) a. nay-ka chinkwu-tul-hanthey-*nun* seys-hanthey X-mas card-lul
     I-nom friend-plural-to-contrast three-to   acc
     ponay-ess-ta.
     send-past-indicative
     'I sent Christmas cards to three friends (contrastive to other groups of
       people, e.g., teachers).'
  b. nay-ka chinkwu-tul-hanthey-*man* seys-hanthey X-mas card-lul
                                 only
     ponay-ess-ta.
     send-past-indicative
     'I sent Christmas cards only to three friends (not to any other groups of
       people).'
  c. nay-ka chinkwu-tul-hanthey seys-hanthey-*na*   X-mas card-lul
                                  as many as
     ponay-ess-ta.
     send-past-indicative
     'I sent Christmas cards to as many as three friends.'

(12) a. nay-ka CHINKWU-TUL-HANTHEY seys-hanthey X-mas card-lul
        ponay-ess-ta.
    'I sent Christmas cards to three friends (contrastive to other groups of
    people).'
    b. nay-ka chinkwu-tul-hanthey SEYS-HANTHEY X-mas card-lul
        ponay-ess-ta.
    'I sent Christmas cards to three friends (contrastive to four of them).'

Next, from the viewpoint of lexical phonology, Y.Cho (1988) presents phonological/morphological evidence for the lexical treatment of all case markers. Here, I will consider only one piece of evidence which clearly shows that non-nominative/accusative case markers should be treated as suffixes. In Korean, there is a productive rule that neutralizes a continuant to a stop only in the syllable coda position, that is, Coda Neutralization (example (13)). However, when the continuant in the stem is syllabified as the onset of the following suffix, it escapes the application of this Coda Neutralization rule (example (14)). This lexical syllabification occurs between the stem and the suffixes (either derivational or inflectional), but never between two morphological words.

(13) a. [k'oč-ilɨm]   -> [k'odirɨm]
        flower-name   'the name of a flower'
    b. [k'oč alɨmtapta]   -> [k'od arɨmdapt'a]
        flower be beautiful   'The flower is beautiful.'
(14) a. [us-ɨm]   -> [usɨm (*[udɨm])
        smile-nominalizer   'smile'
    b. [us-ɨmyənsə]   -> [usɨmyənsə] (*[udɨmyənsə])
        smile-while   'while smiling'

As (15) demonstrates, Coda Neutralization fails to apply between a noun stem and a non-nominative/accusative case marker.

(15) a.[os-ey]   -> [osey] (*[odey])
        clothes-on   'on the clothes'
    b.[k'oč-ɨloputʰə]   -> [k'očɨrobutʰə] (*[k'odɨrobutʰə])
        flower-from   'from the flower'

This clearly shows that these case markers are suffixes attached to the stem in the lexicon, rather than separate morphological words.
    In this section, we have considered two kinds of problems in (10). For an optimal account of quantifier float, we first need some mechanism to distinguish between nominative/accusative case and other case markers, so that we can distinguish between (1-3) and (4-5). Next, on the basis of this dichotomy, we have to figure out why delimiters or stress as exemplified in (11-12) can override the condition in terms of case marking.

3. My Analysis
3.1. Distinction between Structural Case and Semantic Case
    Above, I have argued that both nominative/accusative and other case markers should be inflectional suffixes, without any category change. Assuming this is correct, we still have to find another way to distinguish between the two groups. I propose a distinction in terms of structural case and semantic case (for a similar distinction, see Chomsky 1981, Gerdts and Youn 1988, Cho and Sells 1990). They are characterized in Korean as follows:

(16)   Structural Case   Semantic Case
  a.   structural information   only semantic information
  b.   no case preservation   case preservation
  c.   deletable   no deletion possible
  d.   appears last on the NP   attached at the noun stem

First, nominative and accusative belong to the group of structural case.[7] Both of them (i) carry structural information, as (17) illustrates (nominative: [NP,S], accusative: [NP, VP] in most cases); (ii) are not preserved when they compete with some delimiters, as in (18); (iii) are deletable when the grammatical relations are clear from the context (e.g., in the unmarked word order), as in (19); and (iv) appear as the last suffix in the NP form, as (20) demonstrates.

(17) a. John-i   o-ass-ta.
             nom come-past-indicative
          'John came.'
     b. nay-ka John-ul   manna-ass-ta.
        I-nom   acc meet-past-indicative
          'I met John.'
(18) a. John-to   (*John-ka-to / *John-to-ka) o-ass-ta.
             also   nom-also   also-nom come-past-indicative
          'John also came.'
     b. nay-ka John-to   (*John-lul-to / *John-to-lul)   manna-ass-ta.
                      also   acc-also   also-acc meet-past-indicative
          'I met John, too.'
(19) a. John o-ass-ni?
             come-past-interrogative
          'Did John come?'
     b. na John manna-ass-e.
        I   meet-past-indicative
          'I met John.'
(20) ai-tul-hanthey-man-i/lul
     child-plural-to-only-nom/acc
     'only to children'

In contrast, all the other case markers are semantic in the sense that they (i) carry semantic information, as in (21); (ii) are preserved when they compete with delimiters, as in (22); (iii) are not deletable (example (23)); (iv) are attached to the noun stem as (24) demonstrates.

(21) a. hanthey, eykey : to (goal, experiencer, possessor)
     b. hantheyse : from (source)
     c. ey : to (goal), at (location)
     d. ulo: by (instrumental)
(22) a. John-hanthey-to
                to-also
          'to John too'
     b. John-hantheyse-cocha
                from-even
          'even from John'
(23) a. na-nun ku hakkyo-eyse John-ul manna-ass-e.
        I-topic the school-at   acc meet-past-indicative
          'I met John at the school.'
     b.*na ku hakkyo John manna-ass-e.

(24) ai-tul-hanthey(-man-i)
    child-plural-to(-only-nom)

The second and the third characteristics of the semantic case markers seem to be a natural result of the first one: they should be preserved in any case in order to avoid loss of semantic information.

Then, on the basis of this dichotomy, how can we account for the quantifier float construction?

3.2. Alternative Analysis of Quantifier Float

We may rephrase Gerdts' condition given in (10) in terms of a distinction between structural case and semantic case. However, we still cannot explain how non-nominative/accusative case markers can antecede quantifier phrases when a delimiter is attached to either the noun phrase or the quantifier phrase or when either the noun phrase or the quantifier phrase gets stress as we have observed in (11-12), repeated here as (25-26).

(25) a. nay-ka chinkwu-tul-hanthey-*nun* seys-hanthey X-mas card-lul
        I-nom friend-plural-to-contrast three-to   acc
        ponay-ess-ta.
        send-past-indicative
        'I sent Christmas cards to three friends (contrastive to other groups of
          people, e.g., teachers).'
     b. nay-ka chinkwu-tul-hanthey-*man* seys-hanthey X-mas card-lul
                                  only
        ponay-ess-ta.
        send-past-indicative
        'I sent Christmas cards only to three friends (not to any other groups of
          people).'
     c. nay-ka chinkwu-tul-hanthey seys-hanthey-*na*   X-mas card-lul
                                              as many as
        ponay-ess-ta.
        send-past-indicative
        'I sent Christmas cards to as many as three friends.'
(26) a. nay-ka CHINKWU-TUL-HANTHEY seys-hanthey X-mas card-lul
        ponay-ess-ta.
        'I sent Christmas cards to three friends (contrastive to other groups of
          people).'
     b. nay-ka chinkwu-tul-hanthey SEYS-HANTHEY X-mas card-lul
        ponay-ess-ta.   ·
        'I sent Christmas cards to three friends (contrastive to four of them).'

On the basis of these data, I propose an alternative view of this construction. All the previous studies concern only the antecedent of the quantifier phrase: its grammatical function, case marking, or syntactic category. In contrast, I argue that the relationship between the NP and the quantifier phrase provides us with a simpler and correct generalization. First, I will consider the structure of an NP in Korean and propose two types of semantic information carried by an NP. Second, I will show that a quantifier phrase is an NP. Third, I will propose a condition on the cooccurrence of NPs to the effect of avoiding the redundancy of semantic information. Last, I will show that this rule is independently required in Korean grammar by showing that this quantifier float construction is one instance of so-called multiple case marking constructions in Korean, and that our condition equally applies to all the cases belonging to the construction.

First, let us consider the structure of an NP in Korean. The classification of case markers provides us with a simple account of the structure of an NP (for the use of the term *root*, see Cho and Sells 1990).

(27) Root - Semantic Case - Delimiters - Structural Case

In general, an NP in Korean consists of a root and inflectional suffixes of three classes as illustrated in (27). A root can have only one semantic case (example (28a)) or one structural case (example (28b)). Or a semantic case can be followed by a structural case in some instances as in (28c). Delimiters are particles carrying various kinds of semantic information, for example, contrast (e.g., *nun* 'as for') or scope (e.g., *man* 'only', *to* 'also'). They generally come between a semantic case and a structural case, although some of them (e.g., *nun* 'as for', *to* 'also', *cocha* 'even') override a structural case as in (28d-e).

(28) a. John-hanthey
                to
    b. John-i
                nom
    c. John-hanthey-ka
                to-nom
    d. John-hanthey-man-i
                to-only-nom
            'only to John'
    e. John-hantey-to(-*ka)
                to-also(*-nom)
            'to John too'

Given these considerations, we can say that inflectional suffixes of an NP carry two types of semantic information: one is thematic roles expressed by means of semantic case markers, and the other is expressed by delimiters. Let me call the former Type 1 information, and the latter Type 2 information. Structural case markers do not carry any semantic information. We should note that stress also carries Type 2 information, that is, a contrastive meaning.

Second, let us consider the syntactic category of the quantifier phrase.

(29) a. seys-*hanthey*
        three-to
    b. seys-*kwa* neys
        three-and four
    c. *ku* seys
        the three

The quantifier phrase is clearly an NP: (i) semantic case is attached to it (example (29a)); (ii) it can be coordinated with another noun by means of *kwa* (example (29b)); (iii) it can be modified by a demonstrative adjective (example (29c)). Then, we can say that a typical quantifier float construction as in (1b) is the case where two NPs with the same case occur in a clause.

On the basis of these considerations, I propose a condition on the cooccurrence of NPs (for a similar proposal in Hindi, see T. Mohanan 1990).

(30) NPs carrying exactly and only the same type of semantic information are
        disallowed in a clause, where type of information = Type 1 and Type 2.

The intuition behind this condition is to avoid redundancy of semantic information. This condition guarantees that two NPs in a clause be distinguishable in at least one of the two types of semantic information, that is, Type 1 information carried by semantic case markers and Type 2 carried by delimiters and stress. This condition accounts for (31).

(31) a. haksayng-i   seys-i
        student-nom three-nom
    b.*haksayng-hanthey seys-hanthey
                to   to
    c.*haksayng-hanthey-man seys-hanthey-man
                to-only   to-only
    d. haksayng-hanthey-man seys-hanthey
                to-only   only
    e. haksayng-hanthey seys-hanthey-man
                to   to-only

(31a) is grammatical, since no semantic information is involved. (31b) is ungrammatical, since the two NPs carry exactly and only the same kind of semantic information, that is, Type 1. (31c) is explained in the same way. In (31d), the first NP carries Type 1 and Type 2 information, while the second one carries only Type 1 information. Accordingly, (31d) does not violate (30). (31e) is accounted for in the same way.

Some clarification of this condition is in order. First, this condition cannot be purely phonological, since a series of two NPs with different phonological realizations of one semantic case (example (32a)) or of one delimiter (example (32b)) is also ruled out.

(32) a.*haksayng-eykey seys-hanthey
        student-to   three-to
        'to three students'
    b.*haksayng-eykey-kkaci seys-eykey-cocha
        student-to-even   three-to-even
        'even to three students'

Second, this condition does not apply to NPs belonging to different clauses. In (33), *Mary-eykey* and *ai-tul-eykey* can cooccur since the former belongs to a higher clause.

(33) nay-ka Mary-eykey [John-i   ai-tul-eykey   sathang-ul
     I-nom   to   nom child-plural-to candy-acc
     cwu-ess-ta]-ko   malha-yess-ta.
     give-past-indicative-comp tell-past-indicative
     'I told Mary that John gave candies to children.'

Given these considerations, (30) accounts for all the quantifier float data presented so far: (1-3) are grammatical, since NPs in structural case cooccur; (4-5) are ungrammatical due to a repetition of Type 1 information; (6-8) are explained in any theory assuming empty NPs in the embedded clause, as Gerdts notes;[8] (9) involves only NPs with structural case; and last, (11-12) involve NPs carrying Type 1 information, but they are distinguished in terms of Type 2 information.

Lastly, we observe that this general condition on a series of NPs is equally applicable to all the cases of so-called multiple case marking constructions.[9] (34) demonstrates that multiple nominative/accusative marked NPs can occur in a clause.

(34) a. ku cip-i   cengwon-i nelp-ta.
         the house-nom garden-nom be big-indicative
         'The garden of the house is big.'
     b. na-nun ku cip-ul   cengwon-ul po-ass-ta.
         I-topic the house-acc garden-acc see-past-indicative
         'I saw the garden of the house.'

However, this multiple case marking is not always possible, as in (35).

(35) a. *ku cip-ulo   cengwon-ulo ka-ca.
          the house-to garden-to   go-let's
          'Let's go to the garden of the house.'
     b. *ku cip-uloputhe cengwon-uloputhe naymsay-ka na-nta.
          the house-from garden-from   smell-nom come-indicative
          'A smell is coming from the garden of the house.'
     c. *ku cip-eyse cengwon-eyse party-ka   yelli-nta.
          the house-at garden-at   nom be held-indicative
          'A party is going to be held in the garden of the house.'

If the NPs are marked in the same semantic case, they cannot cooccur in a clause. However, if a delimiter is attached to either of them or if one of the NPs receives contrastive stress, the sentences become grammatical.

(36) a. ku cip-ulo-nun   cengwon-ulo ka-ca.
                     contrast
     b. ku cip-uloputhe-nun   cengwon-uloputhe naymsay-ka na-nta.
                         contrast
     c. ku cip-eyse cengwon-eyse-man party-ka yelli-nta.
                                   only
(37) a. KU CIP-ULO cengwon-ulo ka-ca.
     b. ku cip-uloputhe CENGWON-ULOPUTHE naymsay-ka na-nta.
     c. ku cip-eyse CENGWON-EYSE party-ka yelli-nta.

This contrast between (35) and (36-37) is exactly the same as in the quantifier float construction. Accordingly, the condition in (30) is equally applicable to this case. This is why we have multiple nominative/accusative constructions, rather than multiple dative/ablative constructions.

A closer look reveals that the quantifier float construction is an instance of this multiple case marking construction. First, the two constructions are the same as far as the structure in terms of discourse functions is concerned.

(38) a. ku cip-i   cengwon-i nelp-ta.
         the house-nom garden-nom be big-indicative
         'Speaking of the house, (its) garden is big.'
     b. haksayng-i seys-i   o-ass-ta.
         student-nom three-nom come-past-indicative
         'Speaking of students, three (of them) came.'

The closest translations of the sentences are as given in (38). The first NP is interpreted as a theme, in the sense that it specifies the domain or the universe of discourse with respect to which the predication following it is going to present some relevant information (Dik 1978:132).[10] However, the two constructions differ in the semantic relationship between the two NPs: (i) (38a) shows a whole-part relationship, while (38b) an appositive one (Lee 1989); (ii) the second phrase is an anaphor only in (38b).

These differences result in some differences in word order, as in (39-42). First, within an NP, an NP denoting a part cannot precede one denoting a whole, as in (39b). In contrast, either of them can precede the other one if they are in an appositive relation as in (40).[11] Second, as the quantifier phrase is an anaphor, it should be bound by some NP. Accordingly, (42b) is ruled out. This condition does not hold in non-anaphoric cases such as (41).

(39) a. ku cip   cengwon
       the house garden
     b. *cengwon ku cip
        garden   the house
(40) a. chinkwu   seys (or sey myeng)
       friend   three   three classifier
     b. sey myeng chinkwu
        three classifier friend
(41) a. ku cip-i   cengwon-i nelp-ta.
        the house-nom garden-nom be big-indicative
        'Speaking of the house, (its) garden is big.'
     b. cengwon-i ku cip-i   nelp-ta.
        garden-nom the house-nom be big-indicative
        'Speaking of gardens, (the one of) the house is big.'
(42) a. haksayng-i   sey-i   o-ass-ta.
        student-nom three-nom come-past-indicative
        'Speaking of students, three (of them) came.'
     b. *seys-i   haksayng-i o-ass-ta.
         three-nom student-nom come-past-indicative

On the basis of these discussions, we can safely claim that the condition in (30) is required in Korean grammar for all the cases where multiple NPs occur in a clause. In other words, we do not need a specific condition only for the quantifier float construction, contrary to the claims of previous studies.

4. Conclusion

In this paper, I have proposed an alternative analysis for the quantifier float construction in Korean. In contrast with other works which concentrate on the antecedent NP, a general condition on the redundancy of semantic infomation provides us with a correct generalization.

Notes

*I am grateful to Joan Bresnan, K.P. Mohanan, and Peter Sells for their valuable discussions and comments.

1. Gerdts (1985:59) notes that Korean is distinguished from Japanese in the respect that the floated quantifier can be case marked. As in Gerdts, this paper is concerned only with the cases where the quantifier is case marked.

2. Such forms as *lul uihaye* 'for' or *ey uihay* 'by' should be analyzed as a form consisting of a case marker and adverbial form of a verb, although their meanings are pretty similar to prepositions in English. They change their forms by means of inflection, which is impossible for the other case markers (e.g., *lul uiha-n, ey uiha-n*).

3. For various discussions on grammatical relations in the multiple nominative/accusative constructions, see Park 1981, Chun 1985, Kang 1985, Yoon 1987, among others. As my own data provides a case where the relevant element is clearly neither a subject nor an object, I will not go into details here to figure out which NP is the real subject/object in this construction.

4. Gerdts recently seems to adopt a different view on Korean case marking (Gerdts and Youn 1988:160): she distinguishes between S-case and I-case, which is almost the same as my distinction between structural case and semantic case.

5. There are three views on case markers: (i) to deal with every case marker as an inflectional suffix (J.Cho 1988, Park 1988); (ii) to analyze nominative/accusative as suffixes and others as postpositions (Yoo 1989); (iii) to deal with every case marker as a postposition.

6. The so-called topic marker may be analyzed as a delimiter, in the sense that it adds to a contrastive meaning (Yang 1972). As is predicted, all the topics sanction quantifier float, as in (1).

(1) a. haksayng-hanthey-nun nay-ka seys-hanthey cenhwa-ul ha-yess-ta.
     student-to-topic   I-nom three-to   call-acc   make-past-indicative
     'To students, I made phone calls to three of them.'
   b. chinkwu-hantheyse-nun nay-ka seys-hantheyse cenhwa-ul
     friend-from-topic   I-nom three-from   call-acc
     pat-ass-ta.
     receive-past-indicative
     'From friends, I received phone calls from three of them.'

7. Here, I omit *ui* (genitive), since we are mainly concerned with sentential arguments/adjuncts. It belongs to the structural case: it carries structural information [NP, NP]; is deletable; appears as the last suffix. However, it should be preserved when competing with delimiters.

8. Peter Sells has pointed out to me that we need a more careful characterization of the empty noun. For example, in the framework of Lexical Functional Grammar, this empty noun cannot be accounted for in terms of Functional control due to a case conflict.

9. There are other kinds of multiple nominative constructions, for example, where the first nominative marked NP is interpreted as an experiencer (Gerdts and Youn 1988). They should be distinguished from (31), where a possessor relationship holds between two NPs. However, the condition in (30) is equally applicable to this case.

10. Several studies (Yoon 1987, Heycock and Lee 1989, among others) propose a different pair of terminologies, for example, predication subject and predication, focus and subject. However, I believe that the intuition they attempt to express is the same as mine. I do not think that such terms as topic or focus give the correct generalization, since the first nominative marked NP can be interpreted as a topic or a focus depending on the discourse context. I propose the notion of theme to cover both.

11. Two points are in order: (i) in (39-42), I deliberately select examples where the genitive marker *ui* is deleted to simplify my argument. The examples where *ui* is involved need more explanation; (ii) in (40b), I substitute *seys* 'three' by *sey myeng* 'three person (classifier)'. There is no meaning difference between the two expressions. But for some mysterious reason, *seys chinkwu* 'three friends' is not grammatical.

References

Belletti, Adriana. 1982. On the anaphoric status of the reciprocal construction in Italian. Linguistic Review. 2.2. 101-38.

Cho, Jae Ohk. 1988. Case and coordination in Korean. Korean Linguistics. 5.

Cho, Young-mee Yu. 1988. Phonological evidence for the lexical treatment of Korean suffixes. Paper presented at the LSA annual meeting. New Orleans.

Cho, Young-mee Yu and Peter Sells. 1990. A lexical account of phrasal suffixes in Korean. Ms., Stanford.

Chomsky, Noam. 1981. Lectures on Government and Binding. Dordrecht: Foris Publications.

Chun, Soon Ae. 1985. Possessor ascension for multiple case sentences. Harvard Studies in Korean Linguistics. 1. 30-9.

Dik, Simon. 1978. Functional Grammar. (North-Holland Linguistic Series). Amsterdam: North-Holland Publishing Company.

Gerdts, Donna. 1985. Surface case and grammatical relations in Korean: The evidence from quantifier float. Harvard Studies in Korean Linguistics. 1. 48-61.

Gerdts, Donna and Cheong Youn. 1988. Korean psych constructions: Advancement or retreat. CLS. 24. 155-75.

Heycock, Caroline and Young-Suk Lee. 1989. Subjects and predication in Japanese and Korean. To appear in the Proceedings of the First Southern California Conference on Japanese and Korean Linguistics.

Kang, Young-Se. 1985. Case marking in Korean. Harvard Studies in Korean Linguistics. 1.

Kiparsky, Paul. 1982. Lexical morphology and phonology. Linguistics in the Morning Calm. ed. by I. Yang. Seoul: Hanshin.

Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. Phonology Yearbook. 2.

Lee, Chungmin. 1989. (In)definites, case markers, classifiers, and quantifiers in Korean. Harvard Studies in Korean Linguistics. 3.

Mohanan, K.P. 1982. Lexical Phonology. MIT Diss.

Mohanan, Tara. 1990. Arguments in Hindi. Stanford Diss.

O'Grady, William. 1982. The syntax and semantics of quantifier placement. Linguistics. 20. 519-39.

Park, Byung-Soo. 1981. Double subject revisited. Linguistics in the Morning Calm. ed. by I. Yang. Seoul: Hanshin.

Park, Byung-Soo. 1988. How to handle Korean verbal suffixes in GPSG. Proceedings of the International Circle of Korean Linguistics. 6.

Shibatani, Masayoshi. 1977. Grammatical relations and surface cases. Language. 53.4. 789-809.

Yang, In-Seok. 1972. Korean syntax: Case marking, delimiters, complementation and reflexivization. Seoul: Baikhap Publishing.

Yoo, Beom. 1989. The morphological status of Korean case markers. Working Papers in Linguistics. 37. 28-38. Columbus: Ohio State University.

Yoon, James Hye-Suk. 1987. Some queries concerning the syntax of multiple subject constructions in Korean. Harvard Studies in Korean Linguistics. 2. 138-62.

A Neurophysiological Investigation of *Wh*-Islands[*]

Robert Kluender
University of California, San Diego

1. Introduction

There is a growing body of evidence for the role of lexical semantic factors in what have heretofore been considered purely syntactic processes. Centineo (1986), Van Valin (1987, to appear), Zaenen (1988, 1989), and McClure (1990) have pointed out the role of aspectual verb classes (*Aktionsarten*) in characterizing the class of unaccusative verbs crosslinguistically; Pinker (1989) has proposed a lexical semantic account of the verb classes that participate in various argument structure-changing rules in English (passive, dative shift, locative alternation, and causativization). In this paper I argue that taking a similar approach to the analysis of unbounded dependencies in English provides insight into many poorly understood features of these constructions.

Generative linguistics has traditionally provided structural explanations for processes of extraction in general, and at least since Ross's (1967) original proposals, for island constraints in particular. In Chomsky's (1986a) barriers framework (cf. also Frampton 1990), the standard island constraints are handled by a version of subjacency which equates severity of violation in extraction of arguments with the number of syntactic barriers crossed: 'Our intuitive idea is that movement should become "worse" as more barriers are crossed, the best case being the crossing of zero barriers.' (Chomsky 1986a:28)

At the same time, various researchers over the years have pointed out the role of non-structural factors in processes of extraction. The notion of 'semantic dominance' was first introduced by Erteschik Shir in her dissertation (1977), and further developed in Erteschik Shir and Lappin (1979) and Erteschik Shir (1981); the definition is given in (1).

(1) A constituent *c* of a sentence *S* is dominant in *S* if and only if the speaker intends to direct the hearer's attention to the intension of *c*, by uttering *S*. (Erteschik Shir and Lappin 1979, Erteschik Shir 1981)

Here the intension of a constituent can be intuitively thought of as its semantic content. The hypothesis that results from this is given in (2).

(2) An NP can only be extracted out of clauses which may be interpreted as dominant or out of phrases in which the NP may itself be regarded as dominant. (Erteschik Shir and Lappin 1979)

Additional insight into non-structural factors affecting extraction was provided by Kuno (1976), who originally introduced a thematic constraint on relativization which simply stated that a relative clause must be a statement about its head noun. This idea is expanded upon in Kuno (1987), which provides a more general constraint on extraction, the Topichood Condition.

(3) Only those constituents in a sentence that qualify as the topic of the sentence can undergo extraction processes (i.e. *Wh-Q* Movement, *Wh*-Relative Movement, Topicalization, and *It*-Clefting). (Kuno 1987)

This condition not only bears a certain resemblance to the second half of the disjunction in (2), but is also reminiscent of the first condition of Keenan's (1974) Functional Principle.

(4) (i)   The reference of the argument expression must be determinable *independently* of the meaning or reference of the function symbol.

     (ii) Functions which apply to the argument however may vary with the choice of argument (and so need not be independent of it). (Keenan 1974)

Finally, Ross (1987) has proposed a general account of syntactic prototypes which is at least in spirit not all that different from the barriers proposals.

'The idea here is that it is possible for a sentence to deviate from a prototype, and yet not manifest any drop in acceptability. Losses in viability are cumulative, and only when there have been enough of them for a certain threshold value to be exceeded will the speakers of the language perceive that the sentence is less than perfect.' (Ross 1987:310)

This paper is an attempt to look for correlates of these notions in the processing of extraction structures. It suggests that the effects of subjacency, dominance, topichood, function-argument structure, and prototypicality or viability can to a large extent be subsumed under the notion of 'semantic barrier'. Specifically, the following claims are made.

(5) (a)   Open-class, low-frequency, referentially specific constituents are the best candidates for extraction but simultaneously difficult to extract over ('semantic barriers').

     (b)   Conversely, closed-class, high-frequency, referentially non-specific constituents are relatively easy to extract over.

     (c)   Severity of violation in processes of extraction can at least in part be equated with the number of semantic barriers crossed.

     (d)   The subtle effects of semantic barriers should be reflected in any direct measure of processing load.

The structure of the paper is as follows. After some preliminary discussion, in section 2 I provide evidence for claims (5a) and (5b) in the form of known exceptions to the complex noun phrase, *wh*-island, and coordinate structure constraints in English. The cumulative effects of semantic barriers (5c) are shown in section 3. Finally, in section 4 I propose an experimental means of validating the hypotheses proposed in this paper (5d).

2. The Nature of Semantic Barriers

The notion of semantic barrierhood is intimately tied to the structure of the lexicon and the referential properties of individual lexical items. Psycholinguists traditionally distinguish between two broad lexical classes, namely 'open-class' or 'content' words (nouns, verbs, adjectives, and derived adverbs), and 'closed-class' or 'function' words (articles, conjunctions, prepositions, and pronouns). However,

there is evidence that this categorical distinction has been superimposed on an underlying continuum.

Gentner (1981) has shown for example that in terms of frequency distribution, the pattern exhibited by verbs is intermediate between that of nouns and function words. Nouns are most heavily represented in the low frequency ranges, in classic open-class fashion. Function words are almost exclusively represented in the highest frequency range, the classic closed-class pattern. Verbs are more evenly distributed across all frequency ranges; although they are also represented in the lower frequency ranges, unlike function words, their numbers taper off at these lower frequencies, unlike nouns.

Gradations of class membership exist within lexical categories as well. Prepositions are known to appear in speech errors characteristic of both open-class and closed-class items. Further, Friederici (1985) has shown in a study of the effects of context on word monitoring that prepositions behave more like open-class items when used referentially, i.e. when selected for their meaning, and more like closed-class items when selected merely as part of a verb's subcategorization frame.

Returning to the notion of semantic barrier, the way the preceding facts about the lexicon relate to questions of extraction is as follows: gradations of acceptability in unbounded dependencies often reflect gradations of class membership and referentiality in the constituents of these structures (the extracted element itself, the matrix subject, the matrix predicate, the head noun in a complex noun phrase, the embedded *wh*-phrase or complementizer in an embedded question, and the embedded subject and predicate). Constituents nearer the open-class, low-frequency, referentially specific end of the open- and closed-class continuum (which I will henceforth refer to merely as 'open-class' elements) are optimal candidates for extraction because they are maximally salient. Such elements satisfy the first condition of Keenan's (1974) Functional Principle, which requires the reference of argument expressions to be readily identifiable independent of what follows in the function symbol.

On the other hand, the salience of these same open-class elements makes them semantic barriers when they occur within the function symbol; as Keenan (1974) points out in reference to restrictive relative clauses, in such cases there is more than one independent reference to be made. This has the following effects on the processing of such structures. First, in Kuno's (1976, 1987) terms, it makes it more difficult to determine what the sentence is about, and what should qualify as its topic. Second, it makes filler-gap assignment (Fodor 1978) more difficult by placing greater demands on short-term memory.

Conversely, closed-class, high-frequency, referentially non-specific constituents (henceforth simply 'closed-class' elements) occurring within the function symbol do not produce these effects by virtue of the fact that no independent reference is invoked. Rather, in fulfillment of the second condition of the Functional Principle, they keep the function symbol referentially dependent on, or semantically subordinate to, the argument expression.

If Ross (1987) is correct, then the effects of semantic barriers may not be consciously perceived until enough of them have accumulated in any one structure to collectively draw attention to themselves. As I will show in section 3, this is

generally the case in standard island violations: they contain a number of semantic barriers. On the other hand, exceptions to the various island constraints noted (or more often footnoted) in the generative literature are usually void of semantic barriers. Representative examples are presented in this section, arranged according to the constituent where a barrier can appear. However, crucial distinctions have often been masked in the existing literature by the use of divergent lexical content across examples or by the inclusion of more than one semantic barrier within examples. To avoid these problems, the examples in this section have been set up as minimal pairs wherever possible: paired sentences contrast in one barrier only. As a result, the contrasts are often pretty slim, but only those examples have been included where most people are able to get the relevant contrasts. Since the distinctions are often not even a question mark away from each other, however, I have provided relative markers as guides: the equals sign indicates 'seems about as good as', and the greater-than sign indicates 'seems better than'. Where there is no marker, it is not clear if a contrast exists, though some people get one. Hopefully, when the experimental data are all in it will be possible to talk about these distinctions with greater precision.

2.1. Argument Expressions

Much has been written in recent years about the role of referentiality in processes of *wh*-movement (cf. Pesetsky 1987, Rizzi 1989, and Cinque 1989 & 1990). In particular, it has been demonstrated that the more referentially specific an NP, the better it is as a candidate for extraction. This is shown in (6), taken from Maling and Zaenen (1982:287, fn. 3), based on a similar observation about Italian in Rizzi (1982:70, fn. 5).

(6) (a) *Which article* don't you remember who wrote? >
    (b) *What* don't you remember who wrote? (Maling and Zaenen 1982)

Pesetsky (1987) points out that *which*-N' phrases are more restricted in their set of possible referents than *who, what*, or *how many*-N' phrases (cf. also Rizzi 1989 and Cinque 1989, 1990).

There is a tacit appeal in these accounts to a continuum of referentiality that I would like to make explicit here. Virtually all the island constraint exceptions cited in the literature involve the topicalization, relativization, or clefting of a fully referential NP. When this NP is made increasingly less specific in reference, the following contrasts in extractability emerge.

(7) (a) This is *a paper* that you really need to find someone you can intimidate with.   > (adopted from Chung and McCloskey 1983)
    (b) *Which paper* do you really need to find someone you can intimidate with?
        >
    (c) *How many papers* do you really need to find someone you can intimidate with?   >
    (d) *What* do you really need to find someone you can intimidate with?

(8)   (a)   It's *the Marine Corps* that I know the officers in. (Bolinger 1972) >
     (b)   *Which service branch* do you know the officers in?   >
     (c)   *How many service branches* do you know the officers in? >
     (d)   *What* do you know the officers in?

(9)   (a)   ...*the ponderous formal agreements* that I doubted any Syrian government would be able to sign with the United States and survive. (Goldsmith 1985)
     (b)   *Which ponderous formal agreements* did you doubt any Syrian government would be able to sign with the United States and survive?   >
     (c)   *How many ponderous formal agreements* did you doubt any Syrian government would be able to sign with the United States and survive? >
     (d)   *What* did you doubt any Syrian government would be able to sign with the United States and survive?

(7a) shows relativization out of a relative clause, (8a) clefting out of an NP with a PP complement, and (9a) relativization out of a coordinate structure. When the referential NP is converted to a *which*-N' phrase in (7b), (8b), and (9b), there is a slight deterioration as the set of possible referents is increased. The extraction of a quantified *wh*-phrase in (7c), (8c), and (9c) is somewhat less felicitous, while *what*-substitution in (7d), (8d), and (9d) results in a standard violation. It seems clear that the referential determinability of the argument expression (the first condition of the Functional Principle) plays a crucial role in these examples.[1] The (d) examples show clear frequency and lexical class effects as well: the closed-class interrogative pronoun *what* is much higher in frequency than the corresponding open-class noun phrases.

2.2. Head Nouns in Complex Noun Phrases

In contrast, referential specificity of head nouns impedes extraction out of complex NPs. These well-known facts (as they pertain to picture NPs) are given in (10), and were originally discussed in Chomsky (1973:239, fn. 19).

(10)   (a)   Who did you see pictures of?   >
      (b)   Who did you see *a* picture of?   >
      (c)   Who did you see *the* picture of? >
      (d)   Who did you see *his* picture of? >
      (e)   Who did you see *John's* picture of? (Erteschik Shir and Lappin 1979)

Note that there is an increase both in referential specificity and in degree of open-class membership from (10a) through (10e). (11) shows the same contrast in extraction out of a relative clause.[2]

(11)   (a)   This is the paper that we really need to find *someone* who understands. > (Chung and McCloskey 1983)
      (b)   This is the paper that we really need to find *the guy* who understands. >
      (c)   This is the paper that we really need to find *the linguist* who understands.

Two further observations are in order here. First, virtually all grammatical or marginal extractions out of relative clauses in the literature involve quantified head nouns, which are non-specific in reference. Note also that indefinite pronoun heads such as that in (11a) are closed-class elements. Second, the introduction of a referentially specific head noun in the intermediate clause in (11b) necessitates the

mental identification of an extra referent, thereby increasing the processing load. In Kuno's (1976, 1987) terms, it becomes more difficult to interpret the first head noun as the topic of the sentence. In (11b), it's unclear if the sentence is about *the paper* or *the guy*. This conflict is exacerbated when the head noun is made not only referentially specific but of a lower frequency as well, as shown in (11c).[3]

In sections 2.1. and 2.2. we have seen the following reciprocal relationship: the more referentially specific the argument expression, the better the extraction, and the more referentially specific the head noun of a complex NP out of which extraction occurs, the worse the extraction. This is precisely as predicted by claims (5a) and (5b). In subsequent sections we will explore further cases of this tradeoff between the referentiality of argument expressions and that of expressions figuring in the function symbol.

2.3. Complementizers and Embedded *Wh*-Phrases in *Wh*-Islands

Next we turn to the complementizers and embedded *wh*-specifier phrases of *wh*-islands. In this case we are dealing almost exclusively with what are traditionally considered closed-class elements, yet there are contrasts in the extent to which these elements will license extraction as well. These contrasts are summarized in (12). A general hierarchy of accessibility for extraction out of *wh*-islands is shown on the left; Engdahl's (1983) hierarchy of accessibility for parasitic gap formation is given alongside on the right for comparison.

(12) Accessibility of:

Wh-Islands		Parasitic Gaps			
that	>	manner adverbs	>	(untensed	
if	>	temporal adverbs	>	domains)	>
whether	>	purpose clauses	>		
how	>				
when		*that/than* clauses	>	(tensed	
where	>	*when/because/if* clauses	>	domains)	
why	>				
what/who	>	relative clauses and			
		indirect questions			
				(Engdahl 1983)	

Note that these hierarchies evolved independently of each other. Abstracting away from the difference in finiteness of the embedded clause, which we will return to in section 2.5., there are striking similarities between the two: the ordering of *that* and *if* complementizers, of manner and temporal adverbials, of these two relative to clauses of purpose and causation, and of all of these relative to clauses introduced by *wh*-arguments. I claim that these gradations are due to varying degrees of open-class membership based on referentiality and frequency; those elements which are more referential in character and of a lower frequency are harder to extract over than those with less potential to refer and of a higher frequency.

Turning to the *wh*-island hierarchy on the left, we see a clear increase in degree of open-class membership from top to bottom. Extractions out of *wh*-islands introduced by complementizers are generally pretty good, if not fully acceptable.

These are shown in (13) and (14).

(13) Which medal did Bob doubt   (a) *that* she would win?   >
                                      (b) *if* she would win?   >
                                        (c) *whether* she would win?

(14) Which medal did Bob wonder (a) *if* she would win?   >
                                        (b) *whether* she would win?

Many speakers get the contrasts between the various complementizers in the order given; some don't. However the complementizer *that* is generally assumed to be semantically void of content to the point of being deletable. This is not true of the complementizers *if* and *whether*: *if* refers to a possible state of affairs from among an infinite set of such possible states, while *whether* both historically and synchronically refers to one of only two possible (alternative) states of affairs, and can thus be said to be more referentially specific in character. In addition, there is a marked difference in frequency between the two: *if* has a mean frequency of 2199 in the Francis and Kučera (1982) frequency list for English while *whether* has a mean frequency of only 286.

The adverbials *how, when, where,* and *why* are more referential in character than complementizers, and hence harder to extract over. Example (15) shows minimal contrasts between the complementizer *whether* and the manner adverbial *how.*

(15)   (a)   Which opera singer did you wonder *whether* Mary could stand?
                                     > *how*
    (b)   Which books did he tell the students *whether* they should read?
                                     > *how*

Among the adverbials themselves, the manner adverbial *how* is generally amenable to extraction, the temporal locatives *when* and *where* less so. This is shown in (16), (17), and (18).

(16)   (a)   What do you know *how* to fix? =
    (b)   Who can you remember *how* to imitate? =
    (c)   What did you figure out *how* to finance? (Culicover and Wilkins 1984) >

(17)   (a)   Who did you figure out *when* to visit? =
    (b)   What did you forget *when* to deliver?

(18)   (a)   Who did you forget *where* to seat? =
    (b)   What did you figure out *where* to park?

I have no ready explanation for this difference. However, the contrast in (19) between *when* and *why* is again most likely due to a difference in frequency: a mean frequency of 2333 for *when* versus a mean frequency of 404 for *why.*

(19)   Which book did he tell the class (a) *when* they were going to read? >
                                      (b) *why* they were going to read?

      (20) and (21) contrast *wh*-adverbials with a *wh*-argument.

(20)   (a)   Who did he wonder *when* to give it to? >
    (b)   Who did he wonder *what* to give to?

(21) (a) Who did he wonder *why* he should give anything to? >
     (b) Who did he wonder *what* he should give to?

I claim that this is a difference of class membership and referentiality as well. Interrogative pronouns refer to entities encoded by nominal expressions (i.e. to densely constituted, perceptually or conceptually bounded entities) while *wh*-adverbials refer to the relations predicated of such salient, bounded entities in time and space (cf. Gentner 1981, Langacker 1987). Thus in comparison to interrogative pronouns, *wh*-adverbials exhibit the same intermediate status of class membership that verbs have in comparison to nouns.

To sum up, in this section we have seen how gradations of extractability over the complementizers and *wh*-specifiers of *wh*-islands reflect differences of referentiality and frequency in closed-class elements. These facts closely parallel the contrasts we saw in extractability over the head nouns of complex NPs in the previous section, as well as the contrasts in accessibility of various clause types to parasitic gap formation shown in (12). In the next section we will see similar contrasts in extractability over matrix subjects.

2.4. Embedding Subject

This section shows the curious fact that almost all cases of felicitous extraction out of relative clauses in the literature -- and not just in English -- involve first and second person subjects in the clause immediately superordinate to the relative clause (what I am referring to as the 'embedding' clause). Having a third person subject, as in (22b & c), makes the extraction somewhat worse.

(22) (a) That's one trick that *I've* known a lot of people who've been taken in by.
         > (Chung and McCloskey 1983)
     (b) That's one trick that *he's* known a lot of people who've been taken in by.
         >
     (c) That's one trick that *the attorney's* known a lot of people who've been taken in by.

The crucial factor here appears to be relative cognitive distance from the speech event, sometimes referred to as an empathy hierarchy. The reference of first and second person indexical pronouns is contextually determined and implicitly given, while third person anaphorics are usually dependent for their reference on their referential antecedents. Referential NPs are of course maximally explicit in reference. So there are once again degrees of referential specificity involved in these examples; there is also a clear shift from a closed-class to an open-class subject in (22c). As in the case of referentially specific head nouns (section 2.2.), each independent referential link that needs to be established in the function symbol complicates the processing of the functional expression as a whole.

2.5. Embedded Subject and Predicate

This section demonstrates the well-known fact that it is easier to extract out of *wh*-islands (23) and relative clauses (24) (and also to form parasitic gaps, as indicated in (12) above) when the embedded clause is non-finite and lacks an overt subject.

(23) He told me about a book which I can't figure out...

(a)	whether *to buy* or not.	>	(b)	whether *I should buy* or not.
	how *to read.*	>		how *I should read.*
	where *to obtain.*	>		where *I should obtain.*
	what *to do* about.	>		what *I should do* about.
				(Ross 1967)

(24) (a) This is a paper that we really need to find someone *to intimidate* with. >

(b) This is a paper that we really need to find someone *we can intimidate* with.

It is obvious that a specified subject is more referentially specific than an unspecified one, and hence harder to extract over. The reason is that once again an additional independent reference is required in the function symbol.

Although in the above examples it is impossible to dissociate finite morphology on the verb from the presence of an overt subject, (25) demonstrates that finiteness alone can act as a semantic barrier itself.

(25) (a) That's the kid that I found a book *for you to read* out loud to. >

(b) That's the kid that I found a book *(that) you can read* out loud to.

Finite verb forms (which in English require overt subjects) are referentially more specific than non-finite forms insofar as finiteness relates the time of events referred to in the discourse to the time of the speech event itself. For this reason tense is often considered an indexical expression. This is reflected in the increased morphological complexity of finite verb forms, and in this sense they can be considered semantic barriers in their own right.

2.6. Embedding Predicate

Much attention has been paid over the years to the role of the matrix predicate in processes of extraction. In this section I show that viewing these cases in light of the distinctions presented above, namely in terms of where the embedding predicate falls on the open- vs. closed-class, frequency and referentiality continuum, can help make some sense of why particular verb classes should be more opaque to extraction than others. In addition, I refer to three relevant lines of current research: the study of complex predicates or 'light' verbs (Cattell 1984, Grimshaw and Mester 1988, Kearns 1988, Sells 1989, Dubinsky 1989, Pelletier 1990, Di Sciullo and Rosen 1990), of argument structure-changing rules (Pinker 1989), and of unaccusative verbs (Centineo 1986, Van Valin 1987 & to appear, Zaenen 1988 & 1989, McClure 1990).

Light verbs are typically non-specific in reference, picking up their reference from their NP complements instead. The most common light verbs in English are the high frequency verbs *do, have, give, make,* and *take.* Pinker (1989) points out that these verbs act as little more than slot-fillers or tense-carriers in English, and that the role they perform is often filled by affixes in other languages; several of the researchers listed above have suggested that these verbs have unspecified or only partially specified argument structures. These facts taken together lead to the conclusion that these verbs have a status closer to that of function words, which predicts that they should be easier to extract over than other types of verbs, in

accordance with claim (5b).

Pinker (1989) points out that notions of 'manner' turn up consistently in the definition of verb classes subject to argument-structure changing rules, as do notions of accomplishment, creation, and coming into existence. These latter notions figure prominently in lexical semantic accounts of unaccusativity as well. The four-way classification of verbs into stative, activity, achievement, and accomplishment verbs proposed in Vendler (1967) and refined in Dowty (1979) has been used with great success to characterize the class of unaccusative verbs in Italian, Dutch, Japanese, and English. In every analysis presented so far (see references above) the set of unaccusative verbs has been derived from the non-activity verb classes, namely stative, achievement, and accomplishment verbs; activity verbs never form part of this set.

Van Valin (to appear) points out that what the non-activity verbs have in common is a state predicate in their logical structure. What I will claim here is that verbs built up from state predicates are by nature less referential than verbs with activity predicates. This predicts that, all other things being equal, it should be easier to extract over non-activity verbs than activity verbs.

With this battery of interrelated facts, the behavior of embedding predicates in unbounded dependencies can be seen to be very systematic. First let us look at verbs which take sentential complements. Factive verbs are generally stative or achievement verbs, and thus extraction over such predicates is for the most part unimpaired as predicted. But within the class of factive verbs there are frequency effects: while extractions over high-frequency factives like *know* and *forget* are unproblematic, extractions over low-frequency emotive factives like *rejoice, exult,* and *grieve* are consistently bad.

Manner-of-speaking verbs are invariably activity verbs, which makes them as a class more difficult to extract over. But within the manner-of-speaking verbs there are also frequency effects. Erteschik Shir (1977) divided manner-of-speaking verbs into two groups based on the judgements of acceptability she got for extraction over such predicates. These are given in (26).

(26) *Questionable*: exclaim, grunt, holler, mumble, murmur, mutter, roar, scream, shout, sigh, snort, stammer, wail, whine

Bad: animadvert, coo, croak, dictate, editorialize, eulogize, jeer, lisp, purr, quip, rumble, simper, snarl, transcribe, ululate (Erteschik Shir 1977)

Checking these two groups against the Francis and Kučera (1982) frequency list for English reveals a mean frequency of 21 for the 'questionable' group and a mean frequency of 3 for the 'bad' group.

Bridge verbs are typically stative, in line with the prediction that verbs with state predicates should be easier to extract over. They are also all very high in frequency. Furthermore, compared to the manner-of-speaking verbs, bridge verbs have little referential content. This is demonstrated in (27), where the contrast is brought into relief by the relative lack of referentiality in the extracted element as well.

(27) (a) How angry did Mary *say* that John was? >
 (b) How angry did Mary *say softly* that John was? =
 (c) How angry did Mary *whisper* that John was? (Culicover and Wilkins 1984)

When the embedding bridge verb in (27a) is made more explicitly referential through the addition of a manner adverbial in (27b), it behaves much like the manner-of-speaking verb in (27c) in impeding extraction.[4]

The increased referentiality of the manner-of-speaking verbs implied by their very name is something that I am claiming is true of all activity verbs. The relative lack of referential content of the bridge verbs, on the other hand, is something that I am claiming is true of stative predicates in general. Bridge verbs contribute nothing substantial to the meaning of the sentence other than to signal the introduction of the proposition to which the hearer's attention is directed. This is what makes sentential complements of bridge verbs 'semantically dominant' in Erteschik Shir's terms. In a way the bridge verbs have become little more than slot-fillers or tense-carriers themselves. In fact, the bridge verbs can be said to behave much like the light verbs, whose sole function seems to be the introduction of their semantically dominant NP complements.

Turning now to verbs which take complex NP complements, we can see that frequency does not play much of a role in complex predicates because light verbs are virtually non-referential in the first place and thus easy to extract over. This is true even when low-frequency variants appear in the verbal position, as in (28b).

(28) (a) What did she *make* a complaint that he had done? =
 (b) What did she *lodge/register/voice* a complaint that he had done? >
 (c) What did she *investigate* a complaint that he had done?

Note further that complex predicates are typically statives with verbs of possession in the verbal position (*have a feeling, hold a belief, entertain hopes*) or accomplishments with verbs of creation in the verbal position (*make a claim, lodge a complaint, advance a hypothesis, put forth an idea, start a rumor*). When the verbal position is filled with an activity verb, as in (28c), extraction is impeded.

Analogous facts pertain in extractions out of picture noun phrases. When the embedding predicate is a stative (*have*), achievement (*see, find*), or accomplishment (*draw, develop*) verb, as in (29a), the extraction is fine. When it is an activity verb (*analyze, criticize, discuss*), as in (29b), the extraction deteriorates. Interestingly, this deterioration also occurs with achievement (*lose*) and accomplishment (*destroy*) verbs that are not verbs of creation or coming into existence.

(29)

What did John	(a)	have		(b)	analyze	a picture of?
		see			criticize	
		find	>		discuss	
		draw			lose	
		develop			destroy	

Similarly, the best embedding predicates in exceptions to the relative clause subcase of the CNPC are stative verbs like *be, have,* and *know,* and cognitive

achievement verbs like *see, hear, notice, recognize, find* and *meet*. Virtually all the exceptions to the CNPC cited in the literature are constructed from verbs of these types. When they are replaced by activity verbs, the extractions suffer. This progressive deterioration is shown in (30).

(30) (a) This is a paper that *there* really *must be* someone who understands. >

(b) This is a paper that we really need to *find* someone who understands. >

(c) This is a paper that we really need to *talk to* someone who understands. >

(d) This is a paper that we really need to *emulate* anyone who understands.

(30c) and (30d) both contain activity verbs; note the additional effect of lower frequency in (30d).

In summary, we have seen once again how factors of frequency, referential specificity, and lexical class membership affect the acceptability of extraction -- this time as manifested in the embedding predicate. In addition, we have seen that the two predictions made earlier in this section have been borne out. First, even low-frequency light verbs seem to pose no barrier to extraction due to their status as function words. Second, activity verbs are consistently more difficult to extract over than non-activity verbs, which we suggested is due to the fact that activity verbs are by nature more referentially specific. Finally, the impeding effect of referential specificity in the embedding predicate falls out from the Functional Principle: a referentially specific predicate within the function symbol requires the establishment of an independent reference on the part of the hearer in addition to that of the argument expression, with a correspondingly higher processing cost.

3. Cumulative Effects of Semantic Barriers

On the basis of the discussion in section 2, we are now ready to turn to the third claim (5c) made at the beginning of this paper: the more semantic barriers crossed in an extraction, the worse it becomes. It was pointed out earlier that the effects of semantic barriers may not be consciously perceived until there are enough of them in any given structure to interfere significantly with processing; hence the potential difficulty of perceiving the contrasts in the previous section. However in the following example it is possible to witness the progressive breakdown of a relatively acceptable extraction out of a relative clause as the closed-class constituents of the sentence are successively replaced with constituents exhibiting more open-class characteristics.

(31) (a) This is a paper that we really need to find someone to intimidate with. >

(b) This is a paper that we really need to find someone *we can intimidate with*. >

(c) This is a paper that we really need to find someone *that we can intimidate* with. >

(d) This is a paper that we really need to find someone *who we can intimidate* with. >

(e) This is the paper that we really need to find *the linguist who we intimidated* with. >

(f)   This is the paper that we really need to *razz the linguist who we intimi-dated* with. >

(g)   This is the paper that *the audience* really need to *razz the linguist who we intimidated* with. >

(h)   This is the paper *which the audience* really need to *razz the linguist who we intimidated* with.

In (31b) finite morphology and an overt subject appear in the embedded clause; in (31c) an overt complementizer appears; in (31d) a *wh*-relative pronoun replaces the complementizer; in (31e) the head noun becomes referentially specific, and in (31f) the embedding predicate becomes a low-frequency activity verb. In (31g) the matrix subject switches from a first person indexical to a referentially specific NP, and in (31h) the complementizer is again replaced with a *wh*-relative pronoun.

In (32a) the initial argument expression is made less referentially specific by changing the extraction type from relativization to question formation with a *which*-N' phrase, and in (32b) the extracted element is made even less referentially specific (and more closed-class in nature) by converting it to an interrogative pronoun.

(32) (a)   *Which paper* do the audience really need to razz the linguist who we intimidated with? >

(b)   *What* do the audience really need to razz the linguist who we intimidated with?

The result is relatively uninterpretable compared to what we started out with in (31a), and looks for all intents and purposes like a 'core case' island violation. Note further the extent to which the argument expression in (32b) depends for its reference on that of the function symbol, making the functional expression as a whole unacceptable (Keenan 1974). It is true that we started out with a (perhaps for some marginal) exception to the CNPC, but the same effect can be demonstrated on a perfectly grammatical, garden-variety long-distance extraction. Limitations of space prevent me from doing so here, and so I will leave this as an exercise for the ambitious reader.

4. Experimental Validation of Semantic Barriers

In closing I would like to discuss briefly the experimental research I am currently engaged in to test the fourth claim (5d) made above, namely that the effects of semantic barriers should be reflected in any direct measure of processing load. The specific predictions that this claim gives rise to are the following. First, extracted open-class words (i.e. topicalized, relativized or clefted NPs and *which*-N' constructions) should require more processing than extracted closed-class words (*who* or *what*). Second, open-class elements intervening between filler and gap should require more processing than closed-class elements. Finally, the increased processing load of open-class elements should be detectable even when they occur in otherwise perfectly grammatical constructions, with or without long extraction.

There are three versions of the current experiment, each designed to elicit a different response to unbounded dependencies: reaction time, scalar judgements of acceptability, and evoked response brain potentials. It is this latter measure which I will focus on here. Evoked response potentials, or ERPs, are computer-averaged

waveforms of the EEG synchronized in response to cognitive events and recorded at various electrode sites on the human scalp. Over the past fifteen years various components of this averaged waveform have been shown to be sensitive to different cognitive factors: attention, predictability, anticipation, etc. In a series of studies, Kutas and Hillyard (1980a, 1980b, 1983) showed that the inclusion of semantic anomalies in otherwise semantically and syntactically well-formed sentences of English (e.g. 'A low often brings *kittens* or snow') presented to subjects one word at a time consistently elicit ERPs with a negative wave component peaking at 400 msec. (the 'N400') after the appearance of the incongruous word. This is a very robust effect and is clearly distinguishable from the typical response to other unexpected or surprising task-relevant stimuli (such as showing the target word in a larger type size), namely a positive wave component peaking around 300 msec. (the 'P300') poststimulus.

Significantly, N400 effects have also been observed in a number of studies involving the manipulation of class membership and frequency. The following facts are known about the N400 in relation to these two factors.

(33) (a) The amplitude of the N400 elicited by content words is larger than the N400 response to function words.

(b) The amplitude of the N400 response to content words decreases monotonically as sentences progress.

(c) This decrease in amplitude reflects the influence of semantic but not syntactic context.

(d) Low-frequency content words elicit a larger N400 than high-frequency content words when semantic context is weak or absent. (Van Petten and Kutas, to appear)

These facts suggest that the N400 is an appropriate measure for investigating the role of lexical semantic factors in unbounded dependencies and testing the set of predictions stated at the beginning of this section. The first set of experiments is designed to study the effects of various complementizers and *wh*-phrases on extraction out of *wh*-islands (see section 2.3.). Later I will be looking at specificity effects in the *wh*-specifiers of *wh*-islands and the head nouns of complex noun phrases, as well as at the effects of subject and verb choice in the superordinate clause on extraction out of an embedding.

In conclusion, the facts illustrated in this paper show that degrees of referentiality, frequency, and lexical class membership have similar and related effects across all the major constituents of an unbounded dependency, and that these factors collectively play a major role in determining the acceptability of extraction. These findings support the idea that the examination of syntactic phenomena in light of lexical semantic theories is a profitable research strategy. In particular, the unexpected role of aspectual verb classes in syntactic phenomena as seemingly unrelated as unaccusativity and unbounded dependencies, and of other semantically determined sub-classes of verbs in argument structure-changing rules and processes of extraction, suggests that these same factors may well be found lurking around other areas of syntax as well. The approach to the study of unbounded dependencies taken in this paper also has certain affinities with formal semantic, pragmatic, and syntactic theories of extraction which treat such structures as instances of predication. This is an idea which I am pursuing in concurrent work (Kluender, in

201

preparation). The strong advantage of this approach is that the claims it makes are amenable to experimental validation in the on-line processing of natural language. As such it presents a unique opportunity to subject the findings of linguistic theory to the scrutiny of cognitive neuroscience.

Footnotes

*This research was supported in part by NIH Grant HD22614. I am grateful to Claudia Brugman, Kathy Carey, Sandy Chung, Jeff Elman, Janet Fodor, Helen Goodluck, Mary Hare, Yuki Kuroda, Marta Kutas, George Lakoff, Ron Langacker, Diane Lillo-Martin, Steve Pinker, Steve Poteet, Karen van Hoek, Cyma Van Petten, and members of the PDP/Natural Language Processing group at UCSD for useful comments and suggestions.

[1]The first condition of the Functional Principle also provides a neat account of the following well-known contrast, first noted in Kuno (1973), which does not appear to follow from referential specificity: extraction of an entire prepositional phrase out of an island is generally better than extraction of the prepositional object alone with preposition stranding in the embedding.

(i) the man *to whom* I wonder what to give >
(ii) the man *whom* I wonder what to give *to* (Chomsky 1986b)

Stranding of the preposition in (ii) makes the reference of the argument expression in some sense dependent on the meaning of the function symbol, and as Keenan (1974:306) points out, the extent to which this is true makes the functional expression as a whole that much less acceptable.

[2]Frampton (1990) notes similar contrasts in the formation of parasitic gaps within relative clauses.

(i) Jack, who [*everyone* who likes *t*] visited *t*... >
(ii) Jack, who [*the man* who likes *t*] visited *t*...

[3]Note that examples (i) and (ii) below show relatively felicitous extractions out of NP PP constructions with fully referential head nouns, whereas (iii) is bad.

(i) It's the front door that I have *the key* to. =
(ii) It's Macy's that I know *the clerks* at. >
(iii) It's the front door that I know *the man* at. (Bolinger 1972)

The reason is that in (i) and (ii) the head noun is an 'attribute', in Kuno's (1987) terms, or an 'integral part', in Erteschik Shir's (1981) terms, of the extracted NP, and hence semantically subordinate to it. In terms of the Functional Principle, this means that only one fully independent reference is required. This is not the case in (iii), where the relationship of *the man* and *the door* is more circumstantial, and two independent references are required.

[4]I am indebted to Knud Lambrecht for this observation.

References

Bolinger, Dwight. 1972. *What Did John Keep the Car that was In? Linguistic Inquiry. 3.109-14.

Cattell, Ray. 1984. Composite Predicates in English. (Syntax and Semantics, 17.) Orlando: Academic Press.

Centineo, Giulia. 1986. A Lexical Theory of Auxiliary Selection in Italian. Davis Working Papers in Linguistics. 1.1-35.

Chomsky, Noam. 1973. Conditions on Transformations. A Festschrift for Morris Halle, ed. by Stephen R. Anderson and Paul Kiparsky, 232-86. New York: Holt, Rinehart and Winston.

Chomsky, Noam. 1986a. Barriers. Cambridge: MIT Press

Chomsky, Noam. 1986b. Knowledge of Language: Its Nature, Origin, and Use. New York: Praeger.

Chung, Sandra and McCloskey, James. 1983. On the Interpretation of Certain Island Facts in GPSG. Linguistic Inquiry. 14.704-13.

Cinque, Gugliemo. 1989. 'Long' Wh-movement and Referentiality. paper presented at the Second Princeton Workshop on Comparative Grammar. Princeton.

Cinque, Gugliemo. 1990. Types of A'-Dependencies. Cambridge: MIT Press.

Culicover, Peter W. and Wilkins, Wendy K. 1984. Locality in Linguistic Theory. Orlando: Academic Press.

Di Sciullo, Anna Maria and Rosen, Sara Thomas. 1990. Light and Semi-light Verb Constructions. paper presented at the Fifth Biennial Conference on Grammatical Relations. UCSD. La Jolla.

Dowty, David R. 1979. Word Meaning and Montague Grammar. Dordrecht: Reidel.

Dubinsky, Stanley. 1990. Light Verbs and Predicate Demotion in Japanese. paper presented at the Fifth Biennial Conference on Grammatical Relations. UCSD. La Jolla.

Engdahl, Elisabet. 1983. Parasitic Gaps. Linguistics and Philosophy. 6.5-34.

Erteschik Shir, Nomi. 1977. On the Nature of Island Constraints. Bloomington: Indiana University Linguistics Club.

Erteschik Shir, Nomi and Lappin, Shalom. 1979. Dominance and the Functional Explanation of Island Phenomena. Theoretical Linguistics. 6.41-86.

Erteschik Shir, Nomi. 1981. More on Extractability from Quasi-NPs. Linguistic Inquiry. 12.665-70.

Fodor, Janet Dean. 1978. Parsing Strategies and Constraints on Transformations. Linguistic Inquiry. 9.427-73.

Frampton, John. 1990. Parasitic Gaps and the Theory of Wh-Chains. Linguistic Inquiry. 21.49-77.

Francis, W. Nelson and Kučera, Henry. 1982. Frequency Analysis of English Usage: Lexicon and Grammar. Boston: Houghton Mifflin.

203

Friederici, Angela D. 1985. Levels of Processing and Vocabulary Types: Evidence from On-Line Comprehension in Normals and Agrammatics. Cognition. 19.133-66.

Gentner, Deidre. 1981. Some Interesting Differences Between Verbs and Nouns. Cognition and Brain Theory. 4.161-78.

Goldsmith, John. 1985. A Principled Exception to the Coordinate Structure Constraint. Papers from the General Session at the Twenty-First Regional Meeting of the Chicago Linguistic Society. ed. by William H. Eilfort, Paul D. Kroeber, and Karen L. Peterson. CLS 21, part 1.133-43. Chicago: Chicago Linguistic Society.

Grimshaw, Jane and Mester, Armin. 1988. Light Verbs and θ-Marking. Linguistic Inquiry. 19.205-32.

Kearns, Kate. 1988. Light Verbs in English. ms. MIT. Cambridge.

Keenan, Edward L. 1974. The Functional Principle: Generalizing the Notion of 'Subject Of'. Papers from the Tenth Regional Meeting of the Chicago Linguistic Society, ed. by Michael W. La Galy, Robert A. Fox, and Anthony Bruck. CLS 10.298-309. Chicago: Chicago Linguistic Society.

Kluender, Robert. in preparation. Semantic Barriers to Extraction: A Processing Account of Island Constraints. The Psycholinguistics of Island Constraints, ed. by Helen Goodluck and Michael Rochemont. Dordrecht: Reidel.

Kuno, Susumu. 1973. Constraints on Internal Clauses and Sentential Subjects. Linguistic Inquiry. 4.363-85.

Kuno, Susumu. 1976. Subject, Theme, and the Speaker's Empathy -- A Reexamination of Relativization Phenomena. Subject and Topic, ed. by Charles N. Li, 418-44. New York: Academic Press.

Kuno, Susumu. 1987. Functional Syntax: Anaphora, Discourse and Empathy. Chicago: University of Chicago Press.

Kutas, Marta and Hillyard, Steven A. 1980a. Event-Related Brain Potentials to Semantically Inappropriate and Surprisingly Large Words. Biological Psychology. 11.99-116.

Kutas, Marta and Hillyard, Steven A. 1980b. Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. Science Washington D.C. 207.203-5.

Kutas, Marta and Hillyard, Steven A. 1983. Event-Related Brain Potentials to Grammatical Errors and Semantic Anomalies. Memory and Cognition. 11.539-50.

Langacker, Ronald W. 1987. Foundations of Cognitive Grammar: Theoretical Perspectives. vol. 1. Stanford: Stanford University Press.

Maling, Joan and Zaenen, Annie. 1982. A Phrase Structure Account of Scandinavian Extraction Phenomena. The Nature of Syntactic Representation, ed. by Pauline Jacobson and Geoffrey K. Pullum, 229-82. Dordrecht: Reidel.

McClure, William. 1990. A Lexical Semantic Explanation for Unaccusative Mismatches. paper presented at the Fifth Biennial Conference on Grammatical Relations. UCSD. La Jolla.

Pelletier, Rosanne. 1990. Light Verbs in Telugu: A Clause Union Analysis. paper presented at the Fifth Biennial Conference on Grammatical Relations. UCSD. La Jolla.

Pesetsky, David. 1987. Wh-in-Situ: Movement and Unselective Binding. The Representation of Indefiniteness, ed. by Eric Reuland and Alice G.B. ter Meulen, 98-129. Cambridge: MIT Press.

Pinker, Steven. 1989. Learnability and Cognition. Cambridge: MIT Press.

Rizzi, Luigi. 1982. Violations of the Wh Island Constraint and the Subjacency Condition. Issues in Italian Syntax, ed. by Luigi Rizzi, 49-76. Dordrecht: Foris.

Rizzi, Luigi. 1989. Relativized Minimality. Cambridge: MIT Press.

Ross, John Robert. 1967. Constraints on Variables in Syntax. Bloomington: Indiana University Linguistics Club.

Ross, John Robert. 1987. Islands and Syntactic Prototypes. Papers from the General Session at the Twenty-Third Annual Regional Meeting of the Chicago Linguistic Society, ed. by Barbara Need, Eric Schiller, and Anna Bosch. CLS 23, part 1.309-20. Chicago: Chicago Linguistic Society.

Sells, Peter. 1989. More on Light Verbs and θ-Marking. ms. Stanford.

Van Petten, Cyma and Kutas, Marta. to appear. Influences of Semantic and Syntactic Context on Open and Closed Class Words. Memory and Cognition.

Van Valin, Robert D., Jr. 1987. The Unaccusative Hypothesis vs. Lexical Semantics: Syntactic vs. Semantic Approaches to Verb Classification. Proceedings of the 17th Annual Meeting of the North Eastern Linguistic Society. Amherst.

Van Valin, Robert D., Jr. to appear. Semantic Parameters of Split Intransitivity. Language.

Vendler, Zeno. 1967. Linguistics in Philosophy. Ithaca: Cornell University Press.

Zaenen, Annie. 1988. Unaccusatives in Dutch and the Syntax-Semantics Interface. CSLI Report No. 123. Stanford.

Zaenen, Annie. 1989. Unaccusativity in Dutch: an Integrated Approach. ms. Xerox-PARC/CSLI-Stanford.

Pragmatic Conventions Influencing Children's Use of Causal Constructions in Natural Discourse

Amy Kyratzis, Jiansheng Guo, and Susan Ervin-Tripp
University of California, Berkeley

Children prior to seven or eight years of age appear to have difficulty comprehending and producing expressions with "because" and "so" on laboratory tests (Corrigan, 1980; Kuhn & Phelps, 1976; Piaget, 1928). In particular, children are known to differ from adults in two ways on these tasks: (1) they invoke different explanations of to-be-explained events (Piaget provides the example of six-year-olds believing that clouds move on their own power) and (2) they sometimes use the opposite order of mention of cause and effect. The second variation is indicated when they are given sentences to complete, for example, *the boy fell down because....*, and provide the logical consequents rather than the antecedents of the specified events.

Traditionally, these differences have been attributed to deficits in general cognitive skills not specific to language, such as the "egocentrism" of the pre-concrete-operational child, or the inability of young children to achieve bidirectional mappings between events (i.e., to reverse the temporal order of events). However, the possibility exists that young children's deficits do not reside in their cognitive ability to understand causal relations among events, but rather, reside in their inability to *express* the relations linguistically.

Interestingly, children do not display these differences when they produce causal expressions spontaneously, in natural discourse (French & Nelson, 1985; Hood & Bloom, 1979; McCabe & Peterson, 1985). This pattern of evidence raises the possibility that there is a mismatch between the conventional pragmatic functions for which causal expressions are used in natural discourse, and those tapped on laboratory tests.

What might the conventional pragmatic functions of causal expressions in natural discourse be? One line of research, that of Susan Ervin-Tripp (Ervin-Tripp & Bocaz, 1989), indicates that *temporal* expressions are most often used by young children in formulating direct and indirect commands, that is, for moves to control others' actions. Like temporal expressions, causal expressions may also first be used by children for interpersonal functions such as controlling others' behavior.

In looking at the data base for the present study, consisting of children's spontaneous productions of causal expressions in natural discourse, we observed a basic distinction that could be drawn among three types of causal expressions. The first type realized interpersonal functions, such as controlling others' behavior, as suggested above. This type we termed *Speech Act-Level Causals*. In this type, the reason clause (i.e., the clause with "because" or "so") justifies why something is said, and why the hearer should comply with what is said, rather than explaining the event that is referred to in the matrix clause. The something that is said

usually (though not always) has some control component to it, as in direct and indirect commands, plans, refusals, and claims. There is no real-world connection between the events described in the two clauses, that is, the event described in the "because" clause cannot be understood as stating the real-world cause of the event or situation described in the main clause. Let us consider Examples (1) and (2) below. (The ages of the speakers are indicated in parentheses at the end of the utterance).

(1) Can I look at it, can I , 'cause I'm the guest? (7;06)
(2) Take the gloves off. Because they'll get dirty. (4;07)

In example (1), being the guest does not cause the object's being looked at, but rather, explains why the speaker is indirectly commanding the listener to allow the speaker to look at the object. In (2), getting dirty is not the real-world cause of the gloves being off, but rather, justifies the speaker's directive to take the gloves off, and provides the rationale for the speaker's compliance.

Not all Speech Act-Level causals justify control acts, as the previous two examples might suggest. It is possible for a Speech Act-Level Causal to justify other types of speech acts, such as predictions, questions, and so on. Consider Example (3) below.

(3) That floor is gonna break 'cause somebody lives down there. (4;07)

As in the previous two examples, the Reason Clause in this example does not denote the real-world cause of the floor's breaking. It justifies why the prediction *That floor is gonna break* is being made, that is, why it's relevant in the situation in which the speaker and hearer find themselves. However, the speech act in this case, the prediction, does not carry any control component. Interestingly, this sub-category of Speech Act-Level Causals was infrequent in the present corpus; most of the Speech Act-Level Causals that children produced justified control acts rather than other types of speech acts.

A second type of causal expression that occurred in our corpus we termed *Content-Level Causals.* In such expressions, the event being explained is the event actually referred to in the main clause, and a real-world causality connects the two events described in the two clauses. Let's consider Examples (3) and (4).

(4) He was barking. Because he wanted to get out. (3;08)
(5) I sprained my ankle 'cause I was hitting my father's shoe. (7;05)

In (4), the dog's desire to get out was the real-world cause of its barking. In (5), the child's hitting his father's shoe was the real-world cause of his spraining his ankle. To utilize a distinction put forth by Halliday (1975), Content-level Causals have a mathetic, ideational function, (i.e., they *describe*) in contrast to Speech Act-Level Causals, which realize a pragmatic, instrumental, interpersonal function (i.e., they *persuade* others and *control* their behavior).

A third type of causal expression that occurred in our data base we termed *Epistemic Causals*. In these expressions, as in Speech Act-Level Causals, the event being explained is not the event referred to propositionally in the main clause. Rather, it is a conclusion arrived upon by the speaker, and the event described in the subordinate clause is the speaker's evidence for the conclusion. Consider Examples (6) and (7).

(6)   Oh yeah, I got 42 because the last one's 42. (4;11)
(7)   This is for gardening, because it's fat. (4;11)

In (6), the last one being 42 is not the real-world cause of the speaker's having 42, but rather, constitutes his evidence for concluding that there are 42. Similarly, the tool's being fat in (7) constitutes the child's evidence for concluding that it is for gardening (since he and his father have been having a discussion in which the father claimed that gardening tools tend to be fat or heavy), rather than constituting the real-world cause of the object's being for gardening.

Interestingly, these distinctions among Speech Act-Level, Content-Level, and Epistemic Causals map onto distinctions that have been observed by Sweetser (1984) among types of causal and conditional constructions in adult discourse. According to Sweetser, recognizing the Speech-Act and Epistemic levels of meaning allows for the disambiguation of many adult uses of causals and conditionals. The purpose of the present study was to examine the relative frequency of these three types of causals in *child* discourse. A second purpose was to examine whether these three types have implications for syntactic expression. That is, are different syntactic forms used to realize the three types of causal expression, and how well do young children distinguish among the forms that are appropriate for expressing the three types? The research of Ervin-Tripp and Bocaz (1989), demonstrating that formal differences among children's *temporal* constructions correspond to different pragmatic functions of the constructions, suggests that this may be the case for *causal* constructions as well.

Method

Subjects and Data Base. In order to study these issues, a corpus of archival data which is available on the Discourse Lab system at the University of California at Berkeley was analyzed. This was the 1976 Ervin-Tripp family corpus. The corpus was as follows. Data were collected from seven families, each with two or three children aged between 2;7 and 11;1 (two years, seven months and eleven years, one month). The corpus is composed of natural family interactions videotaped in the home involving naturalistic activities, for example, role play activities such as birthday party preparation and playing tea party, and non-role play activities such as having lunch and making valentines. Interactants included mothers, fathers, peers, and siblings. There were 21 children in the entire corpus.

Coding. All causal expressions with "because" and "so" produced by the children were extracted. Expressions with variations of "because", such as "'cause", were also included. A *causal expression* was defined as an utterance or

succession of utterances that expresses a causal relation between events or between an event and a speech act. There were 303 such expressions produced by children in the entire corpus.

Each expression was coded for the following features: (1) the pragmatic function of the expression, that is, whether it was a Speech Act-Level Causal, Content- Level Causal, or Epistemic Causal; (2) clause order and connective type; and (3) the type of verb used in each clause, that is, whether it was a telic verb, an activity verb, or a stative verb.

Speech Act-Level Causals, Content-Level Causals, and Epistemic Causals were coded according to the criteria referred to earlier, that is, according to whether the "because" or "so" clause in the expression explicated real-world events or justified speech acts or conclusions. In Content-Level and Epistemic Causals, the matrix clause contained an assertion, while in Speech Act-Level Causals, the matrix clause contained interrogatives, as in (1) and (8), direct imperatives, as in (2), indirect imperatives, as in (9), and responses, as in (10).

(8) How do you know that I'm not 8? 'Cause you're only 7. (5;09)
(9) You have to hold onto him because you don't want him to fall. (4;03)
(10) *Child 1*: Why don't *you* call? *Child 2*: No, because the phone's busy. (4;11)

Despite the fact that in indirect imperatives such as (9), the matrix clause has the form of an assertion or declarative, the expression nonetheless fulfills our criteria for a Speech Act-Level Causal. The expression does not describe the relation between an event and its real-world cause, since the event in the main clause is not something that has actually happened or that will definitely happen, but rather, is being negotiated at the moment. Because the expression contains a directive (albeit an indirect one) and its justification, it constitutes a Speech Act-Level Causal.

Clause order and connective type were coded according to the following categories: (1) *B1 Causals* were constructions with *because* in which the Action Clause comes first, Reason Clause, second, in the order of clauses (e.g., *Let's leave for the restaurant now, because I'm hungry.*); (2) *B2 Causals* were constructions with *because* in which the Reason Clause comes first, Action Clause second, in the order of clauses (e.g., *Because I'm hungry, let's leave for the restaurant now.*); (3) *S1 Causals* were constructions with *so* or *so that* in which the Action Clause comes first, Reason Clause, second, in the order of clauses (e.g., *Let's leave for the theater now, so (that) we'll get there on time.*); and (4) *S2 Causals* were constructions with *so* in which the Reason Clause comes first, Action Clause, second, in the order of clauses (e.g., *I'm hungry, so let's leave for the restaurant now.*).

The data were analyzed in terms of Action and Reason clauses, rather than in terms of matrix and subordinate clauses, in order not to exclude S2 Causals from the analysis. Despite the fact that S2 Causals are coordinate constructions and hence do not contain *subordinate* clauses, they depict the same types of events that

are depicted in the Action and Reason Clauses of B1, B2, and S1 Causals. Hence, they should be included in any analysis of causal construction types.

Type of verb in the two clauses was coded according to a coding scheme developed by Vendler (1967). This coding scheme differentiates between stative and dynamic verbs. *Stative Verbs* are verbs that depict enduring states, such as *be, have, know, want.* These situations remain stable unless something happens to change them. Because they depict enduring states, these verbs are not marked by progressive aspect. Dynamic verbs refer to situations which must be maintained by continued input, and are divided by Vendler into two types, telic and atelic (durative) verbs. Atelic or *Activity* Verbs refer to pure action that has no plausible end point, and can go on as long as the actor wishes. Progressives are common, as in *he was walking for five hours.* For such verbs, one can ask, *when did you stop?.. Telic Verbs,* in contrast, are verbs which depict an endpoint. They are divided by Vendler into two types. *Accomplishment Verbs* refer to actions which can take a long time to reach an endpoint, such as *he's writing an article.* The possibility of asking *how long did it take to finish?* shows that these verbs involve an endpoint, or accomplishment. *Achievement Verbs* refer to punctual events like "break", "start", "recognize", and so on. These events also reach an endpoint, but do so rather quickly, such that one cannot ask how long they took, but rather, when they occurred.

Results

For purposes of data analysis, children were divided into three age groups. Group 1 contained children between 2;4 and 3;6 years. Group 2 contained children between 3;7 and 6;6 years. Group 3 contained children between 6;7 and 12;0 years.

The percentage of Causal Function Types, that is, Speech Act-Level, Content-Level, and Epistemic Causals that was produced at each age is given in Table 1.

Table 1 Percentage of Causal Function Types by Age					
	Causal Function Type				
Age	Content	Speech Act-Level	Epistemic	Incomplete	Total
2.4-3.6	0	22.2	0	77.8	100 (18)
3.7-6.6	15.2	55.8	4.6	24.4	100 (197)
6.7-12.0	23.9	59.7	1.5	14.9	100 (67)

As can be seen, more Speech Act-Level Causals than either Content-Level or Epistemic Causals were produced at all three age levels. The youngest age group produced *only* this type. At all three ages, most of the Speech Act-Level Causals that children produced justified control acts, as opposed to other types of speech acts. There seemed to be a small rise in the use of Content-Level Causals between the second and third age groups, suggesting that Content-Level Causals may be an advanced type that increases in frequency with age. Epistemic Causals were very

infrequent, and remained so throughout the age period studied here. This finding is consistent with Slobin and Aksu-Koc's (1982) finding for evidentials indicating that epistemic meanings emerge later than deontic ones. These findings suggest that epistemic meanings are more cognitively complex for young children. Another notable finding was that in the earliest period, the majority of causals produced by children were incomplete. *Incomplete* meant that the child did not produce a complete causal, that is, that the construction lacked either an Action or Reason Clause, or that the causal was a response to a Why-question. The prevalence of this category in young children suggests that they are not well able to produce both clauses of a causal construction, and that there is a sharp rise in this ability after 3 1/2 years.

To conclude from these findings, the primary purpose for which children utilize causals in their own spontaneous constructions is to justify their speech acts (particularly their control acts). Hence, causals for young children appear to be used most often to realize a coercive, interpersonal function, rather than the purely ideational function of explaining events described in propositions. What are the implications of this finding? First, the recognition of the speech-act level meaning of children's causals is important, because, without recognizing this level, many of children's causals would appear illogical or ill-formed. For example, (1), (*Can I look at it, can I, 'cause I'm the guest*), appears illogical unless we realize that the speaker is justifying his indirect directive to the listener and is not providing the real-world cause of the object's being looked at. Similarly, (3), (*That floor is gonna break 'cause somebody lives down there*), appears illogical without realizing that the speaker is justifying why he is predicting that the floor is gonna break -- he does not really think that somebody living down there will actually cause the floor to break in the real world. In a similar way, as noted by Sweetser (1984), recognizing the speech-act and epistemic levels of meaning disambiguates many of *adults'* uses of causals.

A second implication is an ontogenetic one concerning how children learn to use causals. It appears that children first learn causals as realizing one particular pragmatic function, that of justifying speech acts, and only later, with development, acquire a broader range of pragmatic functions for their causal constructions -- their function of explaining events described in propositions and their function of justifying conclusions. As noted previously, children in our corpus used the Speech-Act Level function of causals most frequently in the context of formulating and justifying control acts. Hence, it is not surprising that the Speech Act-Level function of causals emerges earlier ontogenetically, since it is a practical one in terms of getting things accomplished in the child's world. Moreover, a preliminary analysis of the *adults'* uses of causals in this corpus revealed that a vast majority were also Speech Act-Level Causals. This suggests that the Speech Act-Level use of causals is one that is modelled by adults in their speech to children.

A third implication of this finding is a syntactic one. According to Quirk, Greenbaum, Leech, & Svartik (1985), Speech Act-Causals are disjunctive, that is, are more peripheral to their superordinate clause, than are Content-Level Causals, which are adjuncts. According to these authors, the peripheral status of disjuncts

is indicated by the fact that they do not allow a number of syntactic processes to apply to them that are allowed by adjuncts. These processes "reflect a measure of integration within the superordinate clause (p. 1070)". For example, the causal clause of disjunctive constructions lies outside the scope of the negation of the main clause. If you negate the main clause in (2), ("Don't take the gloves off because they'll get dirty"), the scope of the negation is limited to the directive and does not span the reason clause. However, in (3), the scope of the negation does span the reason clause. In other words, "He wasn't barking because he wanted to get out" can be paraphrased "the reason that he was barking was not that he wanted to get out".

It is noteworthy that children first acquire, or are most facile with, constructions which are disjunctive and lie outside of the scope of the sentence, rather than sentence-internal constructions. This finding must be accommodated within any model of language acquisition, since it suggests that sentence-internal relations are not primary, as would be predicted by nativist theories of language acquisition (Chomsky, 1968).

Are there any implications of the functional differences among types of causal constructions for their syntactical realization, that is, for formal features such as clause order or type of connective used? In order to answer this question, we examined the distribution of Clause Order Types by Causal Function, and that analysis is contained in Table 2.

Table 2 — Percentage of Clause Order Types by Causal Function					
Causal Function	**Clause Order Type**				
	B1	**B2**	**S1**	**S2**	**Total**
Content-Level	56.4	1.8	9.1	32.7	100 (55)
Speech Act-Level	60.1	0.6	20.2	19.0	100 (163)
Epistemic	90.0	0	10.0	0	100 (10)

B1 = Action 1st, Reason 2nd (Connective = *Because*)
B2 = Reason 1st, Action 2nd (Connective = *Because*)
S1 = Action 1st, Reason 2nd (Connective = *So*)
S2 = Reason 1st, Action 2nd (Connective = *So*)

It is noteworthy that B1 is always the clause order of choice for constructions with *because*, and that B2 is very infrequently used. This result can be accommodated within Schiffrin's model of the discourse functions of causal expressions. According to Schiffrin (1987), reasons constitute backgrounded events in discourse which support main events, and therefore become subordinated. From this view, one might predict that reason clauses, as supportive material, would come second in the order of clauses. Constructions that conform to this clause order constraint (Action Clause first, Reason Clause second) should be cognitively less complex (from a discourse-processing point of view), and hence, this would explain the preference of young children for B1 over B2 constructions.

The function of the causal construction, that is, whether it was a Content-Level Causal or Speech Act-Level Causal, had implications for whether S1 or S2 was preferred, as can be seen in the interaction in Table 2. S1 use was heightened in Speech Act-Level Causals and S2 was heightened in Content-Level Causals. S1 causals refer to purpose or desired outcomes (consider examples (11) and (12) below), and hence, they may be particularly useful for justifying control acts.

(11) Let's quickly clean it up so your mother won't know. (4;03)
(12) Mom, would you give it to me yellow and white and blue so I can paint? (5;09)

The heightened use of the S2 form in Content-Level Causals may be due to the fact that they maintain the temporal order of events, and do not subordinate the antecedent event that leads up to the main event. Hence, they encode build-up or a logical progression of events, as in (13) and (14).

(13) The fixers fixed the tin man before the witch could fight the tin man, so the tinman had a chance to fight the witch. (4;11)
(14) While the boy was walking along the road, he came upon some beautiful roses and daisies. So, he decided to pick some. (6;11)

For this reason, S2 causals may be particularly appropriate for telling stories or describing sequences of events, hence accounting for their greater use in Content-Level Causals.

Also important to account for is why S1 and S2, both "so" constructions, are used much less frequently by children than B1. One possible account for the infrequent use of S1, which designates *purpose*, is that it constitutes too narrow a range of meaning to be generally useful. In contrast, S2, B1, and B2, all encode a much broader range of semantic meanings, including purpose (as in Example 1), motivating circumstance (as in Example 15), physical cause (as in Example 5), and other meanings.

(15) Can I change Elsa's water because she's thirsty? (6;02)

Concerning S2, the fact that it violates the discourse constraint suggested by Schiffrin's model (i.e., Action Clause, first, Reason Clause, second), may account for its infrequent use by young children. This hypothesis is supported by the fact that the use of S2 causals in our corpus increased with age. If S2 violates a discourse constraint, this may make it more cognitively difficult for young children, hence accounting for its greater infrequency among the younger age groups. In support of this account, Ervin-Tripp and Bocaz (1989) found that control acts used by young children usually took first position, and preceded temporal clauses (i.e., conformed to Action Clause, first, Reason Clause, second).

Schiffrin's claim concerning the backgrounding function of Reason Clauses in causal constructions receives further support from the following analysis of the

contrastive use of different verb types, Telic, Activity, and Stative verbs, in Reason and Action Clauses.

Table 3 Percentage of Verb Types for Action and Reason Clauses				
	Verb Types			
Clause	Telic	Activity	Stative	Total
Action	46.9	21.4	31.6	100 (196)
Reason	30.0	12.5	57.5	100 (257)

As can be seen, Reason Clauses much more frequently than Action Clauses depict enduring states. Concomitantly, Action Clauses much more frequently than Reason Clauses depict dynamic events, particularly telic ones. This makes sense if we consider the events described in Reason Clauses as background, supportive material for the events described in Action Clauses. In keeping with this view, the events in Reason Clauses should be continuous, ongoing contexts which create a backdrop against which dynamic events described in Action Clauses can be understood or made sense of, or which motivate the dynamic events in Action Clauses. Hence, they are likely to be enduring states, as in (16) below, and also as in Examples (1), (4), (6), (7), (8), (9), (10), (11), and (15) above.

(16) Max, every time he saw one, then he had to watch, then they knew it was a nice one, but--whenever he saw a boy one, then those were the bad one, so they told him to watch out. (4;03)

One might argue that, since the Action Clauses in children's causals often contain an imperative (i.e., in Speech Act-Level Causals), it is not surprising that the events contained within them tend to be dynamic, particularly telic ones. Hence (at least in these cases), the foregrounding function of Action clauses need not be invoked to explain verb choice. However, there is no feature other than the backgrounding function of Reason Clauses to explain why *these* clauses tend to contain stative rather than dynamic verbs. For example, the speaker in (15) could have just as easily provided a dynamic Reason Clause event (e.g., *because she spilled hers; because her glass fell over*), but chose instead to designate a stative one. We suggest, in line with Schiffrin's (1987) claim, that this verb choice is motivated by the backgrounding function which Reason Clauses serve in discourse.

To summarize, the results that emerged from this study were as follows: (1) children evidenced a preferred pragmatic function for their causals, that of justifying speech acts (particularly control acts), while they used other functions of causals (explaining real-world events and justifying conclusions) much less frequently; (2) the pragmatic function of children's causals influenced various syntactical aspects of their constructions, such as clause order and type of connective chosen; and (3) the discourse functions of Action and Reason Clauses (foregrounding and backgrounding events, respectively) influenced clause order of the

constructions, as well as relative verb choice in the two clauses. These findings suggest that the pragmatic and discourse-level functions of causal constructions and their constituent parts, Action and Reason clauses, must be taken into account in understanding children's meanings for their causal constructions, as well as the forms which they select to syntactically realize these meanings.

References

Chomsky, N . (1968). *Language and Mind.* New York: Harcourt, Brace, & Jovanovich.

Corrigan, R. (1980). A scalogram analysis of the development of the use and comprehension of "because" in children. *Child Development,* **46,** 195-201.

Ervin-Tripp, S. & Bocaz, A. (1989). *Quickly, before a witch gets me: Children's temporal conjunctions within speech acts.* Berkeley Cognitive Science Report No. 61. Institute of Cognitive Studies, Berkeley, CA.

French, L. A. & Nelson, K. (1985). *Children's acquisition of relational terms: Some ifs, ors, and buts.* New York: Springer-Verlag.

Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the development of language.* London: Edward Arnold.

Hood, L., & Bloom, L. (1979). What, when, and how about why: A longitudinal study of early expressions of causality. *Monographs of the Society for Research in Child Development,* 44(6, Serial No. 181).

Kuhn, D., & Phelps, H. (1976). The development of children's comprehension of causal direction. *Child Development,* 47, 248-51.

McCabe,A. & Peterson, C. (1985). A naturalistic study of the production of causal connectives by children. *Journal of Child Language, 12,* 145-159.

Piaget, J. (1928). *Judgment and reasoning in the child.* Totowa, New Jersey: Littlefield, Adams.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English Language.* London: Longman.

Schiffrin, D. (1987). *Discourse Markers.* Cambridge: Cambridge University Press.

Slobin, D.I. & Aksu-Koc, A. (1982). Tense, aspect, and modality in the use of the Turkish evidential. In P. Hopper (Ed.), *Tense-aspect: Between semantics and pragmatics.* Amsterdam: John Benjamins.

Sweetser, E. E. (1984). *Semantic structure and semantic change: A cognitive linguistic study of modality, perception, speech acts, and logical relations.* University of California at Berkeley Ph.D. Dissertation.

Vendler, Z. (1967). Linguistics in philosophy. Ithaca, NY: Cornell University Press.

"What, me worry?" - 'Mad Magazine sentences' revisited.

Knud Lambrecht
University of Texas at Austin

1. Introduction.[1]

In an important paper dealing with the relationship between form and function in natural language, Akmajian (1984) discusses a special type of exclamative sentence which he calls 'Mad Magazine sentence' (MM) and which is illustrated in (1):

(1)   a. What, me worry?
        b. My boss give me a raise?! (Ha.)
        c. Him wear a tuxedo?! (Sure.)

On the basis of a number of formal similarities between the MM construction type and the class of imperative sentences in English, Akmajian argues that MMs and imperatives may be generated by the same, highly general, phrase structure rule, "with the proviso that pragmatic principles for the use of imperatives will in fact limit imperatives to a subset of the structures in question" (1984:14). He concludes that neither the Mad Magazine sentence type nor in fact imperative sentences have a special status in a syntactic theory and he suggests that a very general, and perhaps universal, 'Formal Sentence-Type Schema' accounts for most pragmatically specialized sentence patterns.

In this paper, I would like to present an alternative analysis of the sentence type illustrated in (1), following the Construction Grammar approach to syntactic analysis laid out in Fillmore, Kay & O'Connor (1988).[2] I will argue that MM sentences differ from imperatives in so many respects, both pragmatically and formally, that they cannot be subsumed under the same type, hence cannot be invoked as evidence in favor of a universal Formal Sentence-Type Schema of the type advocated by Akmajian.

My paper is organized as follows. In Section 2, I will summarize Akmajian's analysis of MMs and his claim concerning their similarity with imperative sentences. In Section 3, I will briefly discuss the German equivalent of the English MM. Unlike its English cousin, the German construction bears no resemblance whatsoever with imperatives. Formally, the German construction is not a sentence but a sentence fragment. In Section 4, I will return to the English MM construction and I will argue that it may be given essentially the same analysis as its German counterpart. Instead of being related to the imperative, the MM construction is related to (but not identical with) a cross-linguistically widely attested type of topic-comment construction.

2. Akmajian's analysis.

Akmajian defines the discourse function of the sentence type in (1) as follows: MMs "are used by speakers to express surprise, disbelief, skepticism, scorn, and so on, at some situation or event" (p. 2). For example (1c) might be used in a context such as (2):

(2) Speaker A: I hear that John may wear a tuxedo to the ball...
    Speaker B: Him wear a tuxedo?! He doesn't even own a clean shirt.

The syntactic properties of MMs are summarized as follows by Akmajian:

(3) A. The subject is accusative:
        i. What! *Her* call me up?! Never.
        ii. What! **She* call me up?! Never.

     B. Tense and modals never appear:
        i. *Him *gets* a job?!
        ii. *Her {will/might/...} call me up?!

     C. Sentential adverbs do not occur:
        i. What! *Her unfortunately lose her job!

     D. Rules such a Topicalization, which move constituents into COMP,
         apply in a restricted way:
        i. What! Us read that trash novel by tomorrow?!
        ii. What! *That trash novel, us read by tomorrow?!

According to Akmajian, properties (3A) - (3C) show that MMs lack an AUX consti-
tuent in the sense of Akmajian, Steele, & Wasow (1979), i.e. a constituent which
contains Tense and Modal (but not *have* and *be*). This accounts not only for (3B),
but properties (3A) and (3C) will follow automatically, because nominative case on
a subject is possible only in tensed expressions and sentential adverbs are banned in
tenseless environments (as e.g. in imperatives). Property (3D) suggests that the con-
struction has no COMP node, hence no landing site for topicalized constituents.

    Akmajian argues that a phrase structure rule for MMs must make the subject
constituent optional, in order to account for cases such as (4) in which the subject
me can be omitted:

(4) Speaker A: Why don't you get a respectable job?
    Speaker B: (Me) get a respectable job! What do you think I am?

He then observes that the second constituent in MMs does not have to be a verb
phrase but can belong to any major syntactic category. In such cases, the copula
seems to be understood. Consider the examples in (5) through (7):

(5) Speaker A: I think Bronsky is such a clever author.
    Speaker B: What! Bronsky (be) *clever*?! Ha.   (NP - AP)

(6) Speaker A: Do you suppose Larry is a doctor by now?
    Speaker B: Larry (be) *a doctor*?! What a laugh.   (NP - NP)

(7) Speaker A: I'm furious that Mary joined the army.
    Speaker B: What! Mary (be) *in the army*?! It can't be.   (NP - PP)

According to Akmajian, MMs which lack the copula can be used to refer both to
unrealized and to realized states, events, and actions, while all MMs with overt
infinitives (whether copula or full verb) have "fixed irrealis interpretation". In such
cases, the states, events, and actions in question, though unrealized, are under the
control of either the subject of the sentence or the speaker.

Akmajian tentatively suggests the phrase structure rule in (8) to generate all forms of MM sentences:

(8) S_{MM} --> (NP) X^n, where X ranges over N, A, V, and P, and n is maximal.

The rule in (8) states that the subject of MMs is optional and that it is followed by a predicate phrase which may consist of any of the major syntactic categories, with the proviso that X^n never contains AUX (i.e. tense and modals) and that it is distinct from S.

Akmajian then points to certain restrictions on MMs in the form of constraints on the subject position. The first restriction is that MMs may not contain dummy subjects. This is shown in examples (9) through (11):

(9) Speaker A: Damn! There's no more beer left.
     Speaker B: What! *There (be) no more beer?!

(10) Speaker A: It's false that the world is flat.
     Speaker B: What! *It (be) false that the world is flat?!

(11) Speaker A: Those clouds make it look like it might rain again.
     Speaker B: What! *It rain again?! Oh no.

The second restriction concerns anaphoric *it*, which is also banned from subject position of MMs, as shown in (12):

(12) Speaker A: At last I see the book. It's on the table.
     Speaker B: Oh? *It (be) on the table?! We must be blind.

Akmajian argues that these constraints can be accounted for in terms of an independently stateable intonational factor, according to which the subject of an MM forms an obligatory intonation center. The phrases which are acceptable subjects in MMs are just those which can also stand alone as single exclamations. He concludes that the simple rule in (8) can be maintained, even though it overgenerates. To rule out unacceptable sequences, a special property of MMs must be stipulated to the effect that the subject phrase must form an intonation center.

Having formally defined the MM sentence type, Akmajian then raises the question whether "MMs form a special syntactic sentence-type with a unique pragmatic function", resulting in "a one-to-one correspondence between a unique formal sentence-type and a unique use in context for that special formal type" (p. 11). He argues that this view of MMs would be mistaken because MMs share important grammatical properties with imperatives. MMs and imperatives, he claims, should in fact not be formally distinguished. He bases his argument on the following parallels:

(13) A. *Subject Properties*: Both MMs and imperatives have optional subjects (both *you leave!* and *Leave!* are possible) and both have subjects which are obligatory intonation centers (cf. *you leave!* vs. **ya leave!*).

     B. *Lack of AUX*: Both MMs and imperatives are characterized by lack of tense and modals (cf. *Be nice!* vs. *Are nice!* or **Must leave!*) and neither accept sentence adverbs of the *fortunately*-type (cf. (3B) above).

C. *Oddities with Perfective HAVE*: Both MMs and imperatives with perfective *have* seem unacceptable (cf. *?*Have finished our homework by 5?!* and *?*Have finished your homework by 5!*).[3]

D. *Topicalization*: COMP-landing-site rules apply in a restricted way in MMs as in imperatives (cf. (3,D)).

E. *Control-Interpretation in Cases with the Copula*: Both in MMs and in imperatives the presence of a copula indicates that the subject of the sentence, or else the speaker, has control over the state, event, or action in question.

F. *Irrealis Interpretation*: Both MMs and imperatives force an irrealis interpretation of the state, event, or action expressed in the sentence.

Akmajian also mentions a number of differences between MMs and imperatives. For example imperatives can only have a restricted set of subjects, such as *you* and *someone*, and they are restricted to verb phrases with overt verbs. These differences, he argues, follow from the pragmatic principles underlying the use of imperatives. In other words, the fact that only a subset of the structures generated by PS rule (8) are useable as imperatives is due to pragmatic, not syntactic reasons, hence does not have to be mentioned in the syntactic rule. In spite of certain idiosyncratic "local" syntactic features, imperatives are not unique from a "global" standpoint, i.e. "the formal sentence type in general is not restricted to an imperative function" (p. 18).[4] According to Akmajian, it makes therefore no sense to posit an 'imperative sentence' type in a formal grammar, since the basic structure of imperatives is also that of MMs. He concludes that a notion such as 'imperative sentence' has no status in a syntactic theory, but only in a theory of speech acts.

Akmajian then raises the issue whether particular clusterings of formal properties, such as those which imperatives and MMs have in common, should be singled out "as constituting significant sentence-types" (p. 18). Akmajian suggests that the answer to this question is "yes" and that the notion 'sentence-type' indeed has theoretical status in formal grammar. However he argues that such sentence-types belong to a highly restricted, and perhaps universal, 'Formal Sentence-Type Schema'. Although he acknowledges the theoretical possibility of a one-to-one form-function fit, for example in such "highly marked" constructions as *Down with the king!* or *Off with his head!*, his main claim is "that something along the lines of the Formal Sentence-Type Schema, based on a small and restricted set of formal parameters, provides the input from formal grammar to the pragmatics" and that across languages "the task will be to specify a set of correspondence principles that relate certain formal sentence-types and certain pragmatic functions" (p. 21).

In the remainder of this paper, I would like to challenge Akmajian's conclusions by proposing an alternative analysis, in which MMs constitute a formal type in its own right, which cannot be explained in terms of more general syntactic and pragmatic properties of the grammar. I will interpret the MM construction as an instance of what Fillmore, Kay & O'Connor call a FORMAL IDIOM, i.e. "a syntactic pattern dedicated to semantic and pragmatic purposes not knowable from its form alone" (1988:505). While I agree with Akmajian that certain formal properties of MMs (as of imperatives) are predictable from, or motivated by, the pragmatic function of the construction in discourse, I will challenge the idea that these properties

can be accounted for in terms of highly general pragmatic principles and may therefore be left unspecified in the syntactic description of the construction. In MMs, as in many other sentence types, discourse function and formal structure are inseparably intertwined. Idiosyncratic pragmatic and formal features must therefore be specified as inherent properties of this particular construction.

3. Mad Magazine sentences in German.

To avoid the language-specific and culture-specific connotations attached to the label 'Mad Magazine Sentence', I will refer to the German type as the 'Incredulity Response Construction' (IRC), a name suggested by Fillmore, Kay, & O'Connor (1988:511). The German IRC is illustrated in (14), which parallels the English examples in (1):

(14) a.   Ich und mir Sorgen machen?   'Me worry?'
         I-NOM and me-DAT worries make-INF

     b.   Mein Chef und mir eine Gehaltserhöhung geben?   'My boss give me a raise?'
         my-NOM-MASC boss and me-DAT a-ACC-FEM raise give-INF

     c.   Der und einen Smoking anziehen?   'Him wear a tuxedo?'
         that.one-NOM and a-ACC-MASC tuxedo put.on-INF

It is not my purpose here to provide a complete analysis of the German IRC. I will simply point to those grammatical properties in (14) which I think are relevant to a proper understanding of MMs in English.

The IRCs in (14) consist of an NP in the nominative case, followed by the conjunction *und*, and a bare infinitival verb phrase, schematically [NP[+nom] *und* VP[+inf]]. As in English, the German construction does not have to contain an overt verb. If the infinitive is the copula *sein* 'to be', it may be omitted, leaving behind a simple NP, AP, or PP. Verbless IRCs are shown in (15) through (17), which correspond to the English examples in (5) through (7):

(15) Bronsky und schlau?!   (*NP - und - AP*)   'Bronsky clever?!'
     Bronsky and clever

(16) Larry und Arzt?!   (*NP - und - NP*)   'Larry a doctor?!'
     Larry and doctor

(17) Mary und in der Armee?!   (*NP - und - PP*)   'Mary in the army?!'
     Mary and in the army

Notice that in (16) the second constituent is a bare N (*Arzt*). This shows that this NP must be a predicate phrase, since common nouns like *Arzt* may appear without a determiner only if they function as complements of a copula (cf. *Larry ist/bleibt/wird Arzt* 'Larry is/remains/becomes a doctor'). The constituent after *und* can thus be described as an infinitival verb phrase whose infinitive may be missing just in case it is the copula *sein*, in which case the remaining phrase has the form of a complement of *sein*. Since I see no motivation for postulating an optional deletion rule which would apply to *sein* but not to any other infinitive, and since to my knowledge there is no conventional way of designating the particular array of

features required, I will represent the second constituent in the IRC in the form of a disjunction (symbolized by /):[5]

(18)   IRC: [NP[+nom] *und*   VP[+inf]' / YP] , where YP is the complement
              of a copula and where Y is {N,A,P}.

In what follows, I will use the symbol 'PredP[+inf]' as a cover term for the disjunction in (18), i.e. I will use the schema   [NP[+nom] *und* PredP[+inf]] instead of the cumbersome schema in (18).

The construction in (18) clearly does not resemble any known sentence type. In particular, it bears no resemblance to the German imperative construction (in German, imperative verb forms are inflected for person and number). What makes the pattern in (18) unique are the following properties. First, the construction contains an initial NP which by its morphological case and its position ought to be a subject but which is not syntactically a subject because subjects cannot be separated from their predicates by a conjunction and because non-tensed verb phrases cannot have subjects in German. Second the occurrence of the conjunction *und* runs counter to generally shared assumptions about the syntax and semantics of conjoined coordinate structures (the two conjuncts belong to different syntactic categories and the conjunction cannot be interpreted as a zeugma). Finally the particular set of syntactic features required to characterize the second conjunct seems to appear nowhere else in the grammar of German. The German IRC is made up of phrasal constituents which are perfectly ordinary with regard to their internal morphosyntactic structure but which are combined with one another in a non-ordinary, construction-specific way. To use a set of terms which are of prime importance in Construction Grammar, while the INTERNAL SYNTAX of the two major constituents is trivial, their EXTERNAL SYNTAX is a unique feature of this particular construction. As I will show below, the IRC does in fact not represent a sentence, but a conventionalized SENTENCE FRAGMENT.

The idiosyncratic syntactic pattern in (18) is directly paired with a specific discourse function. This function is the same as in English, as far as I can tell. To account for this function, it is necessary to modify Akmajian's characterization (cf. ex. (2) and discussion) in one crucial respect, both for German and for English. The expression of "surprise, disbelief, skepticism, scorn, and so on" is not directed at some situation or event, as Akmajian has it, but at the LINGUISTIC EXPRESSION of a situation or event, i.e. a PROPOSITION which was expressed (or contextually implied) in the immediately preceding discourse. More specifically, it is directed at the pairing of a certain argument with a certain predicate in that proposition. I will call the previously expressed (or understood) proposition the 'context proposition' (cf. Fillmore, Kay, & O'Connor 1988), and I will call the sentence expressing this proposition the 'context sentence'. For instance in the English MM in (2), the NP *him* corresponds to the subject argument *John* in Speaker A's context sentence, and the predicate phrase *wear a tuxedo* corresponds to the VP *may wear a tuxedo to the ball* in the context sentence. For an IRC (MM) to be used appropriately, the NP must semantically correspond to an argument (typically, but not necessarily, the subject) in the context proposition, and the infinitival or verbless predicate phrase must semantically correspond to a predicate with which the NP was associated as an argument. (I am using the vague expressions 'semantic correspondence' and

'association of an argument with a predicate' because of the fact that the meaning correspondence between the IRC and the context proposition need not be one of lexical identity, as shown e.g. in (7).)

We are now in a position to understand the semantic-pragmatic motivation for the syntactic structure in (18). First, since the IRC has the pragmatic function of expressing incredulity with respect to some context proposition, it is natural (though of course not logically necessary) that it should present itself as a kind of incredulous QUOTATION of the relevant parts of the context sentence and that this quotation should have interrogative or exclamative form. The difference between the context sentence and the IRC is then the difference between a 'use' form and a 'mention' form of the same or a similar sentence. However what is "mentioned" in the IRC is not the context sentence itself but its subject and its predicate as separate, "non-sentential" constituents. Under this analysis, the morphosyntactic shape of the NP and the PredP in the IRC follows naturally, since nominative and infinitive are the syntactic default (and hence quotation) forms of the categories N and V in German. (In German dictionaries, nouns are listed in the nominative case and verbs in the bare infinitive.) Second, since what prompted the use of the IRC was the controversial pairing of the argument and the predicate in the context proposition, it is also natural (though again not logically necessary) that the IRC should have the syntax of a conjoined coordinate structure whose conjuncts are the argument and the predicate (or part of the predicate) in the context proposition.

Notice that I am characterizing the relationship between the discourse function of the IRC and its syntactic structure as one of MOTIVATION rather than NECESSITY. Different languages do not necessarily have identical structures to express IRCs, as the English MM shows.[6] This notion of functional motivation entails that it is not possible to omit from the syntactic description of the construction those features which are directly explainable in terms of its discourse function (contradicting Akmajian's claim concerning the irrelevance of certain "predictable" formal features for the syntactic description of MMs and imperatives).

The conjunction of NP[+nom] with PredP[+inf] does not constitute a SEN-TENCE, as Akmajian has it for English, nor does it express a regular PROPOSITION, in the sense that it neither asserts nor presupposes the propositional content which is construable by associating the two conjuncts. What then is the grammatical category of the conjoined structure? To answer this question, it is necessary to take another look at the linguistic environment in which the IRC is conventionally embedded. This environment involves not only the preceding context proposition but also the (overt or implied) proposition which FOLLOWS the IRC, and which expresses the speaker's emotional judgment about the context proposition. An example of an IRC involving such a proposition in overt form is shown in (19):

(19) Der und einen Smoking anziehen? Du hast sie wohl nicht alle!
     'Him wear a tuxedo?! You must be crazy!'

Let us call the proposition expressed by *Du hast sie wohl nicht alle* the 'follow-up proposition'. It seems reasonable to assume that the special meaning associated with the IRC is due to the conventional association of the two conjuncts with such a follow-up proposition, since it is this proposition which expresses the doubt, rejection, challenge, etc. about the context proposition. There is nothing in the syntactic

structure NP[+nom] *und* PredP[+inf] itself which would explain such a connotation. I conclude that the follow-up proposition must be included in the description of the IRC, even though its expression via an overt sentence or sentence fragment is optional.

To account for the external syntax of the conjoined structure in the IRC I would like to suggest that this structure forms a complex TOPIC expression, about whose referent the follow-up proposition expresses a COMMENT. For example in (19) the conjunction of the NP *der* with the VP *einen Smoking anziehen* constitutes a topic about which the sentence *Du hast sie wohl nicht alle* expresses a relevant judgment (namely rejection). I propose to represent the structure of (19) as follows:

(20) $_{TOP}$[$_{NP[+nom]}$[der] und $_{VP[+inf]}$[einen S. anziehen]] $_S$[du (...) alle]

In (20), TOP indicates a syntactic position rather than a syntactic category. Syntactically, the TOP phrase shares with topic constituents in other topic-comment constructions the property of being located outside the domain of S (or of S'). The PredP constituent in this TOP phrase is a possible topic expression because it is infinitival, i.e. nominalized. Pragmatically, the TOP phrase is similar to other topics in that the entity designated by it has been evoked in previous discourse and stands in a pragmatic relation of aboutness with the following proposition (see Lambrecht forthcoming:Ch.4). The structure in (20) is formally and pragmatically related to a crosslinguistically widely attested construction which I call the 'Unlinked-Topic Construction' (Lambrecht, op.cit.), in which an NP expressing a previously evoked discourse entity is preposed as a topic to a sentence whose argument positions are not anaphorically related to the NP and whose proposition is pragmatically construed as conveying information about the entity designated by the NP. (21) and (22) are characteristic attested examples of the Unlinked-Topic Construction in spoken English (the sentences in parentheses indicate the discourse context):

(21) (That isn't the typical family anymore.)
The typical family today, the husband and wife both work.

(22) (Talking about planting flowers)
Tulips, you don't have to plant new bulbs every year.

In (21) and (22), the sentences following the commas express comments about the entities designated by the topic NPs *the typical family today* and *tulips*. As in the IRC, the entities designated by the topic expressions were evoked in the immediately preceding discourse. What distinguishes the IRC from other topic constructions, whether linked or unlinked, and what makes it unique, is the fact that in the IRC one of the topic constituents (the NP) designates a discourse entity, while the other (the PredP) designates an attribute of that entity.

The structure in (20) may be generalized into (23), which replaces (18):

(23) IRC: $_{TOP}$[$_{NP[+nom]}$[...] *und* $_{PREDP[+inf]}$[...]] ($_S$[...])

As part of the grammatical information necessary for the appropriate use and correct interpretation of the IRC the following pragmatic conditions must be appended to the structure in (23): (i) the TOP constituent must correspond semantically to a proposition which was evoked in the immediately preceding discourse and in which the designatum of the PredP[+inf] constituent was predicated of the designatum of the

NP[+nom] constituent; and (ii) the proposition expressed by S must be interpretable as a judgment about the designatum of the TOP phrase. As mentioned earlier, the structure in (23) is a FORMAL IDIOM in the sense of Fillmore, Kay & O'Connor (1988), i.e. it is a fully productive syntactic pattern in which familiar syntactic pieces are unfamiliarly arranged and whose semantic and pragmatic properties do not follow from its form alone but must be stated as features of the construction.

4. Mad Magazine sentences revisited.

I believe that the analysis of the German IRC can be extended to the English Mad Magazine type, with one minor restriction concerning the presence of the conjunction *und* in German. If we adopt my interpretation of the IRC as a special kind of topic-comment construction, most if not all of the formal properties observed by Akmajian can be given a uniform and natural explanation. Moreover certain problems in Akmajian's analysis do not arise in the alternative I propose. As for the parallel between MMs and imperatives, it cannot be maintained under my analysis.

I would like to begin by pointing to an important fact concerning the syntax of MMs which Akmajian does not mention. This is the fact that the sequence of the two initial constituents in the MM construction is REVERSIBLE. Parallel to (1) (a), (b), and (c) we also find the versions in (24):

(24) a. What, worry, me?!
     b. Give me a raise, my boss?!
     c. Wear a tuxedo, him?!

As often observed (e.g. Lambrecht 1981, Bresnan & Mchombo 1987, etc.), constituents in TOP position--being extra-clausal non-argument constituents--may be freely ordered with respect to each other. The phenomenon of reversibility illustrated in (24) follows then directly from my analysis of the MM as a TOP constituent. An attested example of an MM with reversed order (pointed out to me by Sue Schmerling) is the following advertisement for a brand of thermal underwear, in which a happily smiling woman utters the sentence in (25):

(25) Cold, me, never.

(25) is an especially illuminating example as it illustrates, in the concisest possible form, not only the fact that the NP and the predicate phrase may be reversed, but also that the follow-up proposition (here *never*) is a natural part of the construction. The property of reversibility alone is sufficient to demonstrate that the MM cannot be a sentential structure generable with the PS rule in (8). Indeed in English a subject NP can never follow a verb phrase (though it may sometimes follow a verb).

Given the reversibility of the two constituents in TOP position, I propose the structure in (26) for the IRC (MM) construction in English. The symbol PredP[+inf] stands for the same set of syntactic features as in German (cf. (18) and discussion):

(26) MM: $_{TOP}[\,_{NP[+acc]}[\,\dots\,]\,,\,_{PredP[+inf]}[\,\dots\,]\,]\quad(\,_{S}[\,\dots\,]\,)\,]$

The structure in (26) is identical to that in (23), except for the difference in NP case and for the fact that the English structure has a comma instead of a conjunction between the two TOP elements (indicating that the two constituents are not ordered

with respect to each other). Like (23), (26) is interpreted as a formal idiom.

Concerning the various formal properties in item (3) pointed out by Akmajian, these are naturally compatible with my proposal. (3A) (accusative case of the "subject") is explained by the fact that in English the accusative form is the default or unmarked case for pronouns in non-argument positions. (This possibility is hinted at by Akmajian in footnote 3 of his paper). Since TOP phrases are non-arguments, they must have accusative case (cf. the case of the topic pronoun in such non-standard topic constructions as *Me I'm waiting*). The lack of a tense feature stated in (3B) is directly motivated by the pragmatics of the IRC. Speaker B's incredulous response does not concern the time at which the situation or event described in the context proposition takes place but rather the abstract predicate-argument structure of that context proposition. As for the non-occurrence of modals like *will* or *might*, I take it to be a purely formal consequence of the fact that modals have no infinitive in English (in German, where modals do have infinitives, they may occur in IRCs). The fact that sentential adverbs like *fortunately* cannot occur (item (3C)) follows from the semantic property of such adverbs as expressions of a speaker's attitude toward a proposition. Since the incredulous response in the IRC is not directed at Speaker A's attitude towards her proposition but at the meaning of the proposition itself, such adverbs may not occur. Finally the constraint against Topicalization (item (3D)) seems to follow from the fact that the pragmatic appropriateness conditions for topicalizing a constituent are incompatible with the situation under which IRCs are used.[7]

Of central importance for Akmajian's analysis are his observations concerning the optional status of the "subject" NP and the infinitival element in the VP (exx. (4) through (7)). In his attempt to make the syntactic description of the MM maximally general and maximally compatible with the structure of imperatives (leading up to the PS rule in (8)), Akmajian does not consider the following crucial fact: the NP[+acc] and the infinitive are in fact mutually dependent on each other. The NP may be missing only if there is no overt infinitive and the infinitive may occur only if there is an NP. Consider first the modified version of (4) in (4'):

(4') Speaker A: I hear you got a respectable job?
Speaker B: *?Get a respectable job! What do you think I am?

Unlike in Akmajian's example (4), the "subject" cannot be omitted in (4'). The difference in acceptability between (4) and (4') is due to the fact that in (4) the context sentence already contains a tenseless verb form. In (4), Speaker B's reply is acceptable only because it happens to be a direct quote of a portion of the context sentence. If the predicate in the context sentence has a tensed verb form, the NP cannot be omitted in the MM. Since the well-formedness of (4) is contingent upon a formal feature of the context sentence, the omissibility of the NP illustrated by (4) is in fact not a formal feature characterizing MMs.

The same observation holds for Akmajian's examples (5) through (7). While it is true that in these examples the copula is optional, this optionality is contingent upon the presence of the "optional" NP. Compare (5) through (7) with the corresponding versions in (5') through (7'), which lack the overt "subject":

(5') Speaker A: I think Bronsky is such a clever author.
    Speaker B: What! *Be clever?! Ha.

(6') Speaker A: Do you suppose Larry is a doctor by now?
    Speaker B: *Be a doctor?! What a laugh.

(7') Speaker A: I'm furious that Mary joined the army.
    Speaker B: What! *Be in the army?! It can't be.

As in (4'), the versions with the overt infinitive in (5') through (7') are unacceptable without the NP[+acc]. While the infinitival copula is omissible, the NP[+acc] constituent is not. This demonstrates that the "subject" of the MM is in fact not optional and that both NP[+acc] and PredP[+inf] are obligatory elements of the MM construction. Any well-formed utterances which lack the NP constituent are therefore not examples of MMs. Consider the utterances in (27), which lack both the NP and the copula:

(27) a. A job?! b. Clever?! c. A doctor?! d. In the army?!

These are indeed possible replies to the context sentences in (4) through (7), but they are not instances of MM constructions. It is true that with the appropriate intonation these utterances can have a force similar to that of MMs. However in such cases the MM interpretation is not an INHERENT FEATURE of the utterance. Unlike MMs, the utterances in (27) could be used for a variety of different pragmatic purposes. For example they could constitute requests by Speaker B for Speaker A to repeat part of her utterance. The MM interpretation arises with necessity only when the predicate constituent cooccurs with an overt NP as part of a complex TOP constituent or, to introduce a concept from Construction Grammar, when the predicate expression and the argument expression are 'in construction' with one another. The utterances in (27) (and the version in (4) without the NP) are simply quotes of portions of a context sentence. MMs on the other hand have an independent existence as GRAMMATICAL CONSTRUCTIONS. I conclude that the schema in (26) expresses the correct generalization.

The last feature to discuss is the constraint against dummy subjects and the non-occurrence of anaphoric *it* (exx. (9) through (12)), which Akmajian explains in terms of the stipulation that the subject phrase in MMs must form an intonation center. While I agree that the NP in TOP must be accented, I do not think that this prosodic feature should be formulated in terms of a phonological constraint on the syntactic structure of the MM. Rather intonation, syntax, semantics, and pragmatics go hand in hand and correlate with each other in our construction. Dummy subjects are banned from MMs both for intonational and for semantic-pragmatic reasons. Given that the purpose of the MM is to challenge a context proposition, it is natural that dummy syntactic elements like *it* and *there*, which do not contribute to the propositional meaning of a sentence, may not occur. As for anaphoric *it* (ex. (12)), I do follow Akmajian's explanation: the non-occurrence of the personal pronoun *it* is due to the fact that this expression cannot stand alone as a single exclamation.

Let me finish with a remark concerning the two formal differences between MMs and imperatives which Akmajian explicitly acknowledges, i.e. the fact that imperatives obligatorily contain a verb and the fact that they may only occur with a restricted set of subjects. Recall that under Akmajian's analysis these formal

differences do not have to be mentioned in the PS rule in (8) because they follow from general pragmatic principles underlying the use of these constructions. While it is no doubt true that there is a good pragmatic reason for imperatives to have the kinds of (optional) subjects they do, given that they are directed at an addressee; and while the overt presence of a verb is certainly pragmatically motivated in imperatives since it would not make much sense to order someone to do something without specifying the action to be performed--I do not think that Akmajian's conclusion concerning the syntactic identity of MMs and imperatives is theoretically justified, even if we ignore the various other factors which make MMs and imperatives dissimilar. From the observation that certain formal features of a construction are directly motivated by its communicative function it does not follow that such features can be left out of the formal description. To say that they can is a bit like saying that in the description of a bicycle one doesn't have to mention the wheels because the presence of wheels follows from the locomotive purpose of the bicycle. If this line of reasoning were correct, we would be justified in saying, for example, that a bicycle is in fact a kind of chair.

5. Conclusion.

Akmajian's idea that the form-function fit is describable as a mapping function between highly general syntactic types and equally general pragmatic principles has undeniable theoretical appeal. However this idea does not seem to provide a realistic picture of the relationship between form and function in natural language. While it is true that a great many syntactic patterns cannot be uniquely paired with specific uses in discourse, I believe that the number of idiosyncratic form-meaning-use correspondences is much greater than assumed in most current formal approaches. With Fillmore, Kay, and O'Connor (1988) I believe that it is impossible to draw a dividing line on principled grounds between "highly marked" (or 'idiomatic') patterns and "unmarked" (or 'regular') patterns (cf. also Lambrecht 1984 for a similar point). The existence and structure of MMs constitute good evidence in favor of a 'constructionist' approach to syntax in which the fundamental unit of syntactic analysis is not the sentence or the sentence-type, but the grammatical construction.

Endnotes

1. I would like to thank Claudia Brugman, Georgia Green, Paul Kay, Ellen Prince, and Sue Schmerling for various helpful comments.
2. For general discussions of the notion of 'grammatical construction' and statements about the framework of Construction Grammar cf. Fillmore 1988, 1989. For syntactic, semantic, and pragmatic analyses of individual grammatical constructions or classes of constructions within the framework of Construction Grammar cf. Fillmore 1985, Fillmore, Kay, & O'Connor 1988, Kay 1990, Lakoff 1987:462-585, Lambrecht 1984, 1986, 1988, and McCawley 1988.
3. Akmajian claims that the unacceptability of perfective *have* in imperatives has no semantic basis, since synonymous sentences with main verb *have*, such as (i) *Have your homework finished by 5!*, are acceptable. This strikes me as an odd argument. Sentence (i) contains a present imperative and refers to a future state of affairs, hence can hardly be cited as evidence against the existence of a semantic (or

pragmatic) constraint against perfective *have* in imperatives.

4. One "local" formal feature of imperatives which cannot be explained in pragmatic terms is the syntax of *do* and *don't*. Akmajian argues that these two expressions can be analyzed as special sentence-initial 'imperative particles' which can be transformationally inserted in front of the structure [(NP) V^n], thus preserving the validity of PS rule (8). Notice however that this insertion rule permits the generation of sentences like **Do you leave the room!*, whose ungrammaticality Akmajian points out in an earlier passage. See also the discussion of imperative *do* and *don't* in Schmerling 1982:202.

5. The representation of the second IRC constituent as a disjunction was suggested to me by Paul Kay.

6. This is shown also by the Latin IRC, which has a subjunctive rather than an infinitival verb form (*Ego tibi irascar!?* 'Me be angry at you!?'). This subjunctive is sometimes referred to as the 'subjunctive of protest' in Latin grammars, showing that Latin grammarians have recognized the IRC as a grammatical category.

7. I have no satisfying explanation for the fact pointed out by Akmajian that topicalizations without an overt subject seem to be acceptable (cf. *That trash novel, read by tomorrow?!*).

References

Akmajian, Adrian. 1984. "Sentence types and the form-function fit." In NLLT, 2, 1. 1-23.

Akmajian, A., S. Steele, & T. Wasow. 1979. "The category AUX in Universal Grammar." L.I. 10,1.

Bresnan, Joan & Sam Mchombo. 1987. "Topic, pronouns, and agreement in Chicheŵa." In Lg, 63. 741-782.

Fillmore, Charles J. 1985. "Syntactic intrusions and the notion of Grammatical Construction." In BLS, vol. 11. 73-86.

--- 1988. "The mechanisms of 'Construction Grammar'." In BLS, vol. 14. 35-55.

--- 1989. *On grammatical constructions*. Ms. University of California, Berkeley.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. "Regularity and idiomaticity in grammatical constructions: The case of 'Let alone'." Language, 64, 3. 501-538.

Kay, Paul. 1990. "Even." To appear in Linguistics and Philosophy.

Lakoff, George. 1987. *Women, fire, and dangerous things*. Chicago: Chicago University Press.

Lambrecht, Knud. 1981. *Topic, antitopic, and verb-agreement in non-standard French*. Amsterdam: John Benjamins.

--- 1984. "Formulaicity, frame semantics, and pragmatics in German binomial expressions." In Lg, 60:4. 753-796.

--- 1986. "Pragmatically motivated syntax: Presentational cleft constructions in spoken French." In CLS, vol. 22. Papers from the Parasession on Pragmatics and Grammatical Theory. 115-126.

--- 1988. "There was a farmer had a dog: Syntactic Amalgams revisited." BLS, vol. 14. 319-339.

--- Forthcoming. *Information structure and sentence form. The pragmatics of syntax in spoken French.* Cambridge: Cambridge University Press.

McCawley, James D. 1988. "The Comparative Conditional Construction in English, German, and Chinese." In BLS, vol. 14. 176-187.

Schmerling, Susan. 1982. "How imperatives are special, and how they aren't." In CLS, vol. 18. Papers from the Parasession on Nondeclaratives. 202-218.

MENTAL EXPERIENCE VERBS IN MODERN GREEK:
A COGNITIVE EXPLANATION FOR ACTIVE VERSUS MIDDLE VOICE[1]

Linda Manney
University of California, San Diego

0. Introduction

Middle voice is often characterized as indicating that the subject or its interests are affected in some way (Lyons, 1968; Barber, 1975; Gonda, 1975; Klaiman, 1988). Descriptive grammars of Modern Greek, while presenting a slightly more complex picture, nonetheless identify affectedness of subject as a primary meaning encoded by middle voice (Mackridge, 1987; Tzartzanos, 1989). In Modern Greek, mental experience verbs occur with an affected human experiencer subject. Because mental experience verbs in Greek can occur in both active and middle voice, some notion in addition to subject affectedness is necessary in order to account for middle voice in this semantic class.

The present paper explains these phenomena in terms of specific components of prototypical transitivity. The proposed analysis draws on current cognitive theories (Langacker, 1987a; to appear) which elaborate aspects of Hopper and Thompson's (1980) model of event structure and transitivity. I assume that transitivity is a gradient property of clauses, in which the prototypical transitive event involves a transfer of energy from a source (initial endpoint) to an inactive patient-goal (terminal endpoint). It is shown that mental experience events are encoded by middle as opposed to active voice verbs if the conceptualized event instantiates a schema with either a defocused initial or terminal endpoint.

1. Voice Marking in Modern Greek

All Greek verbs, in both finite and non-finite form, are obligatorily marked for active or middle voice. Voice is indicated by verbal suffixes which also encode tense, aspect, and modality. With respect to voice marking, Greek verbs can be grouped into three main classes. The first consists of verbs which occur only in active voice, and in fact have no corresponding middle forms. There is also a very large set of Greek verbs which have both active and middle forms. Some of these verb pairs will be treated in later sections, so I will simply say here that middle morphology encodes a variety of related meanings which can be motivated in particular contexts. Finally, a large number of Greek verbs, approximately one third, occur only in middle voice and in fact have no corresponding active forms (Dr. Roderick Beaton, Modern Greek Department, Kings College, University of London, p.c.). These verbs have traditionally been termed "deponent."[2]

With respect to their arguments, middle voice verbs, both deponents and those with active counterparts, can occur in morphosyntactically transitive as well as intransitive clauses. In the present work, a clause is morphosyntactically transitive only if it contains a nominative case marked subject and an accusative or sentential object. Morphosyntactically intransitive clauses include those with oblique objects (a preposition + an accusative or genitive case object), genitive objects, or no objects. Middle voice verbs can occur with accusative, sentential, genitive, and oblique

objects, as well as with no object, and therefore cut across transitivity which is defined strictly on the basis of morphosyntax.

2. Theoretical Overview

Numerous studies have shown that transitivity is a gradient property of clauses, comprised of a set of related meaning components such as agency, telicity, and object affectedness. (Lakoff, 1977; Hopper and Thompson, 1980; Langacker, to appear). Thus, a prototypical transitive event involves, among other things, a complete transfer between two distinct participants, one of which is an agent and thus willfully initiates the transfer, the other of which is a patient and thus is affected by the agent's action. Hopper (1985) further notes the correlation between object affectedness, referentiality, and transitivity. An affected object is one which is pre-existing at the time of the event designated by a verb and thus capable of being impacted, as opposed to an effected object, which is created during the process designated by a verb. It is claimed that affectedness of an independently existing object may be the most important parameter in determining transitivity. Likewise, non-distinctness of participants is an important parameter in determining departures from canonical transitivity (Langacker and Munro, 1975; Kemmer, 1988).

Cognitive grammar maintains that degree of transitivity is related to construal (Langacker, 1987b; to appear), since the same objective event can be conceptualized in a number of different ways, giving rise to different linguistic constructions associated with the same event. The types of participants selected for linguistic coding therefore reflect the more abstract conceptual organization of the clause.

The present analysis relies on the notion of transitivity as a basic conceptual phenomenon which manifests itself in clause structure to varying degrees. In particular, it is shown that middle versus active voice mental experience events in Modern Greek are consistently distinguished according to less versus greater transitivity, respectively.

3. Active versus Middle Voice Marking on Verbs of Mental Experience[3]

3.1 Overview

In what follows, I explain voice marking on verbs of mental experience in terms of prototypical event structures. I maintain that mental experience events are of two types. The first, Initiative Mental Experience, designates an active experiencer who directs attention to a mental object, and as such, instantiates an extension from the prototypical transitive event discussed above. Morphosyntactically, mental events of this type are typically transitive, i.e. occur with a nominative subject and an accusative or sentential object. Middle voice Initiative Mental Experience is semantically differentiated from the active in that the middle designates an event with an internal, less distinct object (defocused terminal endpoint), as opposed to the active, which designates an external object that exists apart from the experiencer's mental contact with it (focused terminal endpoint). Thus, while the active verbs portray a mental transfer between two distinct participants, the middle verbs, on the other hand, designate an event initiated by an active participant who makes mental contact with a non-distinct second participant.

The second type of mental event, Less Initiative Mental Experience, designates a less active experiencer who is affected by a participant or an event. In the majority of cases, middle versus active voice events differ in that the former focuses on an experiencer's resulting mental state, whereas the latter highlights the cause of a mental state. The middle marked event is typically (but not always) morphosyntactically intransitive: it occurs with a nominative subject which designates an affected entity, and an oblique object or an adverbial clause which identifies the source of the affect. The oblique object in this class of events is therefore not a canonical patient-object, since it is construed as an affecting, rather than an affected entity.

3.2 Initiative Mental Experience: Directed Mental Activity

In this section I will discuss verbs of perception and verbs analogous to perception, both of which are predominantly active voice, and verbs of thinking, which are predominantly middle voice. I claim that both classes of verbs represent extensions from the prototypical transitive event; the middle verbs, however, are farther removed from the transitive prototype. The relevant distinction is that of internal versus external objects, where internal objects correspond to less transitive events, and external objects correspond to events with a higher degree of transitivity.

3.2.1 Active Voice: Verbs of Perception

In Modern Greek, verbs of perception, unlike verbs of cognition or verbs of thinking, are almost all exclusively active voice verbs. The verbs which denote the five senses, for example, are active constructions. Consider the following.

1. vlépis   to   peDĺ ekĺ péra
   2sg=see=ACT ACC child there around
   'Do you see the child over there?'

2. akúo   tin   musikĺ tis   yitónisas mu
   1sg=hear=ACT ACC music GEN neighbor GEN
   'I can hear my neighbor's music.'

3. áNize   to   prósopo tu
   3sg=touch=ACT ACC face   GEN
   'S/he touched his face.'

4. mirĺzo   móno skórDa edó péra
   3sg=smell=ACT only garlic   here around
   'I can smell nothing but garlic around here.'

5. Den borĺ   na   kataláve   tin yévsi tu   faitó
   NEG 3sg=able SUBJ 3sg=understand=SUBJ=ACT ACC taste GEN food
   'He can't taste his food.'

Other verbs which designate more active perceptual activity, i.e., *kitázo* 'to look at,' *paratiró* 'to notice, to observe,' *Diakríno* 'to distinguish,' are also active voice with no corresponding middle forms. Since most other sub-classes of mental experience

verbs to be considered are widely comprised of middle voice verbs, this pattern requires further comment.

In the model of event structure presented above, events are more highly transitive to the extent to which they designate an energy transfer from an energy source to an energy sink. The object in a prototypical transitive construction is totally distinct from the subject, and lies downstream from the flow of energy. It has been pointed out (Langacker, to appear) that verbs which designate a perception event are analogous to prototypical transitive events in which energy expenditure occurs, in that perception is conceived as involving directed activity on the part of the perceiver toward the object of perception. Although the domain is no longer that of physical energy, the focus of attention on an object at the end of a perceptual path is seen as analogous to an energy transfer which affects an energy sink, and thus approximates the transitive prototype. The active voice marking on verbs of perception in Modern Greek therefore finds a natural explanation in terms of prototypical event structure.

The next two active voice mental verbs to be discussed are those which designate the quintessential emotions, *aGapó* 'to love,' and *misó* 'to hate.' I am claiming that unlike most verbs of emotion to be discussed below, *aGapó* and *misó* construe processes in which an active entity directs attention to an external object and therefore represent extensions from the transitive prototype of event structure. They are exemplified in 6.-7. and 8.-9., respectively.

6. aGapái   poli̧ tin   yinéka tu
   3sg=love=ACT a lot ACC wife   GEN
   'He loves his wife very much.'

7. to   aGóri aGapái   poli̧ tin   pi̧isi tu   eli̧ti
   NOM   boy   3sg=love=ACT a lot ACC poetry GEN Elitis
   'The young man loves Elytis' poetry.'

8. o   xri̧stos misi̧   tus   ratsistés
   NOM xristos 3sg=hate=ACT ACC racists
   'Christos hates racists.'

9. ?o   xri̧stos misi̧   tin   fasistiki̧ téxni
   NOM xristos 3sg=hate=ACT ACC fascist   art
   'Christos hates fascist art.'

Both *aGapó* and *misó* require accusative case nominal direct objects; *misó* furthermore prefers animate direct objects. Unlike most middle verbs of emotional response, *aGapó/misó* cannot take clausal complements. Thus, 10. and 11. below are unacceptable in Greek.

10. *to   aGóri aGapái   na   xorépsi
    NOM   boy   3sg=love=ACT SUBJ 3sg=dance=SUBJ=ACT
    'The boy loves to dance.'

11. *o   stratiótis misÍ   na   afÍsi   tin   yinéka tu
    NOM soldier 3sg=hate=ACT SUBJ 3sg=leave=SUBJ=ACT ACC   wife   GEN
    'The soldier hates to leave his wife.'

The verbs *aGapó* and *misó* designate an emotional response directed toward another entity. Their experiencer subjects are thought of as agent-like because they are construed as active beings who focus attention on a distinct second entity. As such, *aGapó* and *misó* parallel verbs of perception. Both types of active verbs construe mental activity which originates with an active experiencer and is directed toward an external entity. Thus, *aGapó* and *misó* designate mental processes which are conceived of as analogous to prototypical transitive events: in both types of events a transfer occurs between an active primary participant and a less active secondary participant.

3.2.2 Middle Voice: Verbs of Thinking

Of the several verbs of thinking activity in Modern Greek, almost all are middle voice, and many of these are deponents. Five verbs of thinking are illustrated in 12. - 16.

12. o   eNlimatÍas analoyÍstike   to   éNlima tu
    NOM criminal 3sg=contemplate=MID/D ACC crime   GEN
    'The criminal thought deeply about his crime.'

13. sképtome   eséna   káTe méra
    1sg=think=MID/D 2sg=ACC   each   day
    'I think about you every day.'

14. fandázome   ti   zoÍ mu Déka xrónya apó sÍmera
    1sg=imagine=MID/D ACC life GEN ten   years   from now
    'I'm imagining my life ten years from now.'

15. skarfÍstike   ena   kólpo
    3sg=think up=MID/D   ACC   trick
    'S/he thought up a trick.'

16. anarotiTÍke   yatÍ Ítane   anixtÍ i   pórta
    3sg=wonder=MID/D   why   3sg=be=MID/D open NOM door
    'S/he wondered why the door was open.'

Unlike perception events, in which the object of perception exists apart from the experiencer's perception of it, verbs of thinking typically involve an experiencer which establishes mental contact with an entity construed as internal to the mental realm. For example, in sentence 12. above the verb *analoyÍstike* designates thinking about a situation or event which is far removed from the experiencer's immediate circumstances. The object of mental contact is not directly observable and is construed as internal to the experiencer. The same can be said of mental objects depicted in sentences 13. and 14: *sképtome*, 'I think,' and *fantázome*, 'I imagine,'

respectively, designate events in which a mental object is construed to be intimately connected to the experiencer's thought processes, although the objects may or may not exist independently. The scene depicted in sentence 15., on the other hand, involves a mental object which is effected by an experiencer, since it comes into being as a result of the process designated by the verb *skarflstike*. Finally, in 16., the mental object of the verb *anarottTike* designates a potential situation or event, i.e. that which is unknown or unrealized. As such, the object has no independent existence apart from the experiencer's thinking of it.

The above examples suggest that the active versus middle voice marking on verbs of perception and thinking, respectively, is semantically motivated. In perception events, the objects of perception typically exist separately from the experiencer, as opposed to objects of thought, which are construed as more intimately connected to the experiencer. Therefore, I maintain that the existence of an external object motivates the active voice in perception events, whereas the internal and often effected object motivates the middle voice in verbs which designate thinking events.

3.3. Less Initiative Mental Experience: Emotional Response

In this section, I discuss mental experience events in which an experiencer is construed passively, unlike the middle events discussed above. The middle marked events to be discussed here designate experiencers which undergo, rather than initiate, experience. In most cases, the middle marked verb construes a psycho-emotional state induced within a sentient being by an external force, where the affected sentient being is a nominative-case-marked experiencer subject, and the stimulus is a prepositional object or a complement clause. The main group of verbs discussed depict emotional response and emotive experience.

Less initiative mental experience is comprised of at least two different types of events. In the first, by far the more common, active versus middle voice corresponds to causation of a state versus resulting state, respectively. The argument put forward here is that middle verbs in this class differ fundamentally from middle verbs with initiative experiencers: they are not extensions from the transitive prototype in which an energy source or its analog is the starting point of the event. What I am claiming is that verbs which designate a one-participant induced state are conceptually basic as compared to those designating the corresponding two-participant event in which both a causer and an experiencer are focal participants. Therefore, of these two possibilities for encoding emotion events, the unmarked verbal construction consists of a verb stem which identifies a particular state, plus the middle morpheme which invokes the notion of a single participant's induced state. On the other hand, the active counterpart, in those cases where one exists, encodes a more complex situation, typically one in which an agent brings about a particular response in another individual.

In the second type of less initiative mental experience, both active and middle events select an inactive experiencer as the primary clausal participant. Here, the active/middle distinction corresponds to a focused versus defocused endpoint along an abstract scale which corresponds to energy flow in a prototypical transitive event.

3.3.1. Active Voice versus Middle Voice: Causation versus State

There are several middle/active pairs in which the middle marked event designates an emotional response, and the active, the causation of an emotional response. Examples 17. and 18. below illustrate two such pairs: the middle constructions in 17a. and 18a. depict an experiencer which is in a particular emotional state; the active sentences in 17b. and 18b. incorporate the agent responsible for the corresponding emotional response in another individual. Most middle verbs in this class require explicit mention of the stimulus, which is typically encoded as an oblique object or as a complement clause, and less frequently as an accusative object.

17a. stenaxoryéme   ya   tin   iyĭa   tu
    1sg=worry=MID/A PREP ACC health GEN
    'I'm worried about his health.'

17b. i   iyĭa tu   me   stenaxorĭ
    NOM health GEN ACC 3sg=worry=ACT
    'His health worries me.'

18a. anastatónete   káTe forá pu vlépi   tus bátsus
    3sg=upset=MID/A each time CMP 3sg=see=ACT ACC police
    'S/he gets upset every time s/he sees the police.'

18b. to   na   di   tus bátsus tin anastatóni
    NOM SUBJ 3sg=see=SUBJ=ACT ACC police ACC 3sg=upset=ACT
    'Seeing the police upsets her.'

In the most basic realization of an emotional event, which is an emotional state, the most prominent participant is an experiencer. As the conceptualization of the event increases in complexity and focuses on responsible rather than affected entities, voice marking reflects this change, resulting in an active voice structure which has an agentive participant as subject.

Sentences 19. and 20. below illustrate that not all middle marked verbs of emotion have active counterparts. However, the existence of deponents is natural and expected according to the present analysis, which claims that the middle marked stative event is the more basic of the two types of events.

19. drépete   ya   tin ftoxĭa tis
    3sg=be ashamed=MID/D PREP ACC poverty GEN
    'She's ashamed of her poverty.'

20. lipáme   pu ĭne   árosti i   mána su
    1sg=be sorry=MID/D CMP 3sg=be=MID/D sick   NOM mom GEN
    'I'm sorry that your mother is sick.'

Notice that the middle constructions in 19. and 20. are formally analogous to those in 17a. and 18a., respectively, since the verbs in both cases occur with non-affected

prepositional objects or sentential complements rather than canonical accusative objects.

In sentences 21. and 22. below, middle and active voice also correlate with basic state versus causation of a state. However, in each respective pair, the middle marked induced state and the active marked causative designate a different type of mental experience undergone by an experiencer.

21a. enDiaférete   ya   ta   provlímata ton   erGatón
     3sg=interest=MID/A PREP ACC problems   GEN workers
     'He is interested in the workers' problems.'

21b. ta   provlímata ton   erGatón ton   enDiaférun
     NOM problems   GEN workers ACC 3pl=interest=ACT
     'The workers' problems interest him .'

22a. apelpízome   me   ton   ánDra   mu
     1sg=lose hope=MID/A PREP ACC husband GEN
     'I'm despairing over my husband.'

22b. me   apelpízi   o   ánDras   mu
     ACC 3sg=deprive of hope=ACT NOM husband GEN
     'My husband really upsets me.'

The middle construction in 21a. means that the entity designated as subject was interested in and personally involved with the workers' problems. It might be said of a foreman who truly cares about the welfare of the workers. Sentence 21b., on the other hand, is not understood in the same way at all. The active verb implies that the person interested was not in any way personally involved with the workers. It might be said of a sociologist who is studying the labor force or by a journalist who is preparing a story, but it would be an inappropriately made statement under the circumstances described in 21a. Likewise, in 22b., the active verb implies a less profound emotional experience as compared to the middle counterpart in 22a. The difference in quality of experience as reflected in the middle versus active marked events is a result of selecting the affected rather than the responsible entity as subject. When an experiencer is the primary clausal participant, the prominence afforded subject position enhances the affected state of the experiencer. On the other hand, when a cause is selected as subject, focus is shifted away from the affected entity, thereby diminishing the effect of the emotional experience.

3.3.2 Active versus Middle Voice: Endpoints

In the next type of middle/active pair to be discussed, both verbs encode stimulus-induced states in which the subject is a non-initiative experiencer. The middle and active events differ, however, in terms of the experiencer reaching or not reaching an endpoint along an abstract scale analogous to a path with an origin and a destination. Consider the following examples.

23a. anéxome   to   póno
     1sg=suffer=MID/A ACC pain
     'I'm tolerating the pain okay.'

23b. andéxo   to   póno
     1sg=suffer=ACT ACC pain
     'I'm feeling extreme pain.'

24a. Den   anéxome   to   póno
     NEG   1sg=suffer=MID/A ACC pain
     'I don't feel any pain.'

24b. Den   andéxo   to   póno
     NEG   1sg=suffer=ACT ACC pain
     'I can't bear the pain.'

In 23.-24., the middle versus the active structures are distinguished according to whether the experiencer does or does not reach her/his threshold of tolerance for pain. The middle structures in 23a. and 24a. mean that the experiencer feels a low level of pain, or none at all, as compared to the respective active structures, which mean that the experiencer reaches an endpoint, i.e. threshold for pain. The active structures are therefore more highly transitive, since they designate an event with an endpoint, as opposed to the middle structures, which designate an experiencer who does not reach an endpoint.[4] Note that in both the active and middle marked events, this abstract endpoint does not correspond to the morphosyntactic direct object, *to póno*, 'ART=ACC pain=ACC,' but rather reflects a point on a more abstract scale which is construed as analogous to a final state in a prototypical transitive interaction.

Further support for this analysis is provided in 25a.-b. below.

25a. Den   andéxi   na   vlépi   tin   yinéka tu
     NEG   3sg=suffer=ACT SUBJ 3sg=see=SUBJ=ACT ACC wife   GEN

     na   kléi
     SUBJ 3sg=cry=SUBJ=ACT
     'He can't bear to see his wife cry.'
     (He feels such strong emotion that he wants to cry.)

25b. Den   anéxete   na   vlépi   tin   yinéka tu
     NEG   3sg=suffer=MID/A SUBJ 3sg=see=SUBJ=ACT ACC wife   GEN

     na   kléi
     SUBJ 3sg=cry=SUBJ=ACT
     'He can't stand to see his wife cry.'
     (He doesn't tolerate her crying very well--he gets angry or nervous.)

In the active structure illustrated in 25a., the subject nominal is construed as having

reached an endpoint with respect to understanding another person's feelings, since the implication is that he suffers the same pain his wife feels when she cries. In 25b., on the other hand, no such implication is made. The mental event construed by the middle verb is that of a negative response toward his wife's crying. No rapport is established, as compared to the active structure, in which the endpoint of establishing rapport is reached.

The next verb to be discussed, *paTéno* 'to suffer, to undergo,' differs from the pair illustrated above in two ways. First of all, it is an active voice verb with no middle counterpart. It encodes a stimulus-induced emotive state in which a nominative case marked affected experiencer is the subject, and the stimulus is often an accusative case marked nominal. *paTéno* also differs from constructions illustrated above in that it is the maximally unmarked verb in Greek to designate negative experience. Thus, unlike middle marked verbs of less initiative experience, *paTéno* can designate both physical and emotional states. Consider the following.

26. i   mitéra   tu   épaTe   katáTlipsi
    NOM mother GEN 3sg=suffer=ACT depression=ACC
    'His mother went through a depression.'

27. tin   épaTa
    3sg=ACC 1sg=suffer=ACT
    'I got mugged, robbed, assaulted, etc.'

28. i   yeoryí épaTan   meGáli   zimyá
    NOM farmers 3pl=suffer=ACT big=ACC loss=ACC
    'The farmers experienced a great loss.' (i.e. There was a crop failure.)

Although the events illustrated in 26.-28. all single out a non-initiative primary participant, sentences 27. and 28. designate resulting states which emphasize physical rather than mental effect. I would like to suggest that this ability to construe both the physical and external as well as the mental and internal motivates the voice marking on the active verb *paTéno,* as opposed to middle verbs illustrated above which can only designate a mental state. Thus, just as external versus internal objects correspond to greater versus less transitivity in initiative mental experience events (e.g. 1.-5. versus 12.-16.), external versus internal effect on an experiencer corresponds to greater versus reduced transitivity in less initiative mental experience.

4. Conclusion

Voice and transitivity in Modern Greek is a tremendously complex system which requires a theory of language sufficiently flexible to accommodate subtle meaning differences encoded by voice marking, yet rigorous enough to explain why particular constructions do not occur. The present work thus attempts to account for a small subset of typical Greek middle constructions using a cognitive theory of language. I have explained active versus middle voice marking within the semantic domain of mental experience by appealing to the notion of transitivity as a gradient property of clauses. I have shown that mental events in Modern Greek are of two

types, Initiative and Less Initiative Mental Experience. In both cases, middle events are consistently less transitive than their active voice counterparts, and in some cases, the middle events are also less marked. These differences, which elude a strictly morphosyntactic account of voice, find a natural explanation when viewed in terms of a cognitively based theory of language.

Notes

[1]I would like to express my deepest appreciation to the following native speakers of Greek for their patience, kindness, and ability to clarify numerous aspects of voice marking in Modern Greek: Yiórgos Kalogerás, Fótis Zaháraglou, and Elizabeth Kotzakídou-Pace. I would also like to express my gratitude to Ron Langacker for his very careful readings of the numerous shapes and forms this paper has taken. Finally, I would like to thank the following people for their helpful comments on the present work: Kathy Carey, Mary Hare, Yiórgos Kalogerás, Ricardo Maldonado, Teenie Matlock, Jo Rubba, Karen Van Hoek, and Maura Velazquez.

[2]I will also use the term 'deponent,' but I use it strictly to refer to the morphosyntactic form of a verb. Therefore, a verb is deponent only if there is no active form built from the same verb stem. According to my definition, then, a verb like *sónome,* 'to run out, to be used up,' is not a deponent, because it has an active form *sóno,* 'to save, to rescue,' although semantically the two are not obviously related. For the purpose of this paper, then, only verbs like *sképtome,* 'to think,' or *lipáme,* 'to be sorry,' are deponents, because neither of these have active voice counterparts.

[3]The following non-standard orthographic conventions have been used in the present work for typographical ease: /D/ = voiced interdental fricative, /T/ = voiceless interdental fricative, /G/ = voiced velar fricative, and /N/ = velar nasal. Voice marking is indicated as follows. MID/D marks a middle voice deponent, i.e., a middle marked verb with no corresponding active counterpart; MID/A marks a middle voice verb with a corresponding active form; ACT marks an active voice verb.

[4]There are also middle/active pairs with non-experiencer subjects that can be distinguished according to whether or not an endpoint along some abstract path is reached or not. Consider the following.

A1. askí   tin   DikiGorikí
    3sg=practice=ACT ACC   law
    'S/he is a lawyer.'

A2. askíte   stin   DikiGorikí
    3sg=practice=MID/A PREP=ACC law
    'S/he is a legal apprentice (but will soon become a lawyer).'

The abstract analog of an endpoint or goal construed in A1. and A2. is the professional recognition and status one receives after having completed all necessary training and study required of lawyers. In the active structure, the person spoken of is a lawyer and therefore has reached this endpoint. The middle structure, on the other hand, construes an agent who has not yet become a lawyer but probably will soon.

Therefore, the middle event, as compared to the active, portrays an agent who has not yet reached the endpoint of professional recognition and thus is construed as less transitive.

References

Barber, E.J.W., 1975. "Voice - Beyond the Passive," *Proceedings of the 1st Annual Meeting of the Berkeley Linguistics Society.* 16-25.

Gonda, J., 1975. "Reflections on the Indo-European Middle, I and II," in *Indo-European Linguistics, Vol. I.* 107-163. Leiden: E.J. Brill.

Hopper, Paul, 1985. "Causes and Effects," *Proceedings from the 21st Regional Meeting of the Chicago Linguistic Society.* 67-89.

Hopper, Paul, and Sandra Thompson, 1980. "Transitivity in Grammar and Discourse," *Language, 56:2.251-299.*

Kemmer, Suzanne, 1988. *The Middle Voice: A Typological and Diachronic Study.* Stanford University Ph.D. Dissertation, Stanford, CA.

Klaiman, M.H., 1988. "Affectedness and Control: A Typological Study of Voice Systems," in *Passive and Voice, Typological Studies in Language, Vol. 16,* 25-83, ed. by Masayoshi Shibatani. Amsterdam/Philadelphia: John Benjamin.

Lakoff, George, 1977. "Linguistic Gestalts," *Proceedings from the 13th Regional Meeting of the Chicago Linguistic Society.* 236-287.

Langacker, Ronald W., 1987a. *Foundations of Cognitive Grammar, Vol. 1.* Stanford: Stanford University Press.

_____, 1987b. "Transitivity, Case, and Grammatical Relations: A Cognitive Grammar Prospectus." Linguistic Agency, University of Duisburg.

_____, to appear. *Foundations of Cognitive Grammar, Vol. 2.* Stanford: Stanford University Press.

_____, and Pamela Munro, 1975. "Passives and Their Meaning," *Language* 51:4.789-830.

Lyons, John, 1968. *Introduction to Theoretical Linguistics.* London: Cambridge University Press.

Mackridge, Peter, 1987. *The Modern Greek Language.* London/New York: Oxford University Press.

Shibatani, Masayoshi, 1985. "Passives and Related Constructions," *Language* 61:4.821-848.

Tzartzanos, Achilleas, 1989. *Neoelliniki Sintaxis tis Koinis Dimotikis, (Modern Demotic Greek Syntax)* Thessaloniki: Ekdotikos Oikos Adelfon Kiriakidi.

Van Oosten, Jeanne, 1986. *The Nature of Subjects, Topics, and Agents: A Cognitive Explanation.* University of California, Berkeley Ph.D. Dissertation, Berkeley, CA.

Speaker-Hearer Asymmetry as a Factor in Language Evolution: A Functional Explanation For Formal Principles Of Grammar*

Frederick J. Newmeyer
University of Washington

Many assume that the positing of an innate universal grammar (UG) forfeits any possibility of a functional explanation for why grammars have the properties that they do. Bates and MacWhinney (1979), for example, contrast the 'functionalist' model, in which the nature of grammars is determined and the child's acquisition of grammar is guided 'by the pragmatic and semantic structure of communication interacting with the performance constraints of the speech channel', with the 'autonomous syntax view', which posits 'abstract categories [and] innate clues about the range of possible human grammars' (1979: 168). It is clear that Bates and McWhinney see these positions as incompatible.

There is a way, however, that autonomous syntax and functionalism can be reconciled. For example, a grammatical principle might have become encoded in our genes by virtue of its being so successful in facilitating communication that the survival possibilities of those possessing it were enhanced. Thus a functional explanation would hold at the evolutionary level.

In this paper, I do in fact propose an explanation based on evolutionary pressure for certain principles of UG. In particular, I raise and attempt to answer the question of why a principle of grammar might become biologized.

It has frequently been suggested that many aspects of grammars have a plausible functional motivation, from predominant word order possibilities of predicates and arguments within clauses (Tomlin 1986) to the types of categories and rules out of which grammars are constructed (Creider 1979) to the principles that constrain their operation (Givón 1979). To the best of my knowledge, however, what has never been observed before is that those grammatical phenomena whose explanation is most convincingly attributed to some principle of UG tend to be those whose functional grounding is asymmetrical between speaker and hearer.

Let us begin with the UG principle of Subjacency, which in English is responsible for violations of grammaticality such as (1a-b)[1]:

(1) a. *What$_i$ do you wonder where John put ___$_i$?
  b. *What$_i$ do you believe the claim that John ate ___$_i$?

Various scholars have pointed to a functional grounding for Subjacency (Givón 1979, Berwick & Weinberg 1984, Frazier 1985a). As they note, violations of this principle tend to be structures which create problems for the hearer in matching the displaced *wh*-element with its coindexed gap. However, it is rarely pointed out that Subjacency performs no particular service for the speaker, whose 'easiest' task would simply be to '*wh*' any Noun Phrase regardless of its subcategorized position in the structure. Hence, Subjacency exhibits a functional asymmetry.

The same point can be made with respect to Principle A of the Binding Theory (Chomsky 1981), for short 'Anaphor Binding'. Violations of this principle are exhibited in (2):

(2) a. *John$_i$ told Mary to help himself$_i$.
  b. *John$_i$ thinks that himself$_i$ should be nominated.

While Anaphor Binding may help the hearer more efficiently to pair anaphoric elements and their antecedents, it seemingly complicates matters for the speaker, who, of course, is fully aware of the identity of the intended referent and is thus forced to make a 'personally'

unnecessary grammatical distinction. Anaphor Binding is therefore functionally asymmetrical as well.

Analogous points can be made with respect to the Case Filter and the ECP (see also Chomsky 1981), though space constraints prevent an elaboration in this paper.

The tendency for innate constraints to exhibit a functional asymmetry is a natural consequence, I believe, of evolutionary pressure for language to serve as an ever more efficient medium of communication. In cases where the ease of the speaker and the needs of the hearer were in direct conflict, an obvious solution presented itself -- to bypass directly the push-pull between speakers' demands and hearers' demands by incorporating those constraints necessary to the hearer directly into the innate language faculty itself. Thus arose the principles of UG, allowing a stable innate core to language, immune to the functional exigencies of the moment.

It may seem at first blush a bit ironic that function-based factors should lead to an innate UG, but a moment's reflection should dispel the irony. If we agree with the functionalist thesis that the ability to communicate by spoken language is a paramount human attribute, and played a major role in the survival and development of the species, then we would *expect* anything that facilitates this process to become biologized. If Subjacency really does ease communicators' burdens, then wouldn't its biologization have conferred an advantage to the species?

There was no evolutionary pressure, however, to biologize what aided speaker and hearer equally. There would hardly be any benefit in encoding in our genes some linguistic principle that the path of least effort would lead both participants in a discourse to follow anyway.

One might wonder whether the existence of parsing principles poses a dilemma for my hypothesis. Clearly, such principles are innate -- we do not 'learn' how to parse sentences. But then, pursuing the logic of the argument, if both speaker and hearer benefit from parsing principles, what caused them to become innate; if, on the other hand, such principles are functionally asymmetrical, then why were they not encoded in UG?

The answer, I believe, is that most, if not all, parsing principles are rooted in memory abilities, and are therefore essentially extensions of a preexisting faculty; that is, parsing principles were the inevitable consequence of the memory faculty adapting itself to the newly-emergent language faculty.

Let us consider by way of example the parsing principle Early Immediate Constituents (EIC), proposed in Hawkins (to appear). EIC asserts that orderings of words are preferred which enable the parser to recognize all IC's of some mother node as rapidly as possible. Thus EIC explains (among other things) why in V-O languages there is a tendency for heavy constituents to occur at the right margins of their verb phrases; why, for example, (3b) and (4b) are preferred to (3a) and (4a) respectively:

(3) a. ?I consider everybody who agrees with me and my disciples about the nature of the cosmos to be smart.
  b. I consider (to be) smart everybody who agrees with me and my disciples about the nature of the cosmos.
(4) a. ?I met the twenty-three people who I had taken Astronomy 201 with last semester in the park.
  b. I met in the park the twenty-three people who I had taken Astronomy 201 with last semester.

As Hawkins notes, 'one of the most fundamental tasks that the syntax module performs, evidently with quite remarkable speed and efficiency, is to group words together into the hierarchically organized phrases of the linguist's constituent structure representation' (ms. p. 6). EIC is instrumental to this task, abetting the speaker *as well as* the hearer in the task of organizing -- as well as perceiving -- rapid speech.

All of the foregoing remarks depend, of course, upon Subjacency and Anaphor Binding being UG-principles. If they (or the effects that led to their postulation) are to be

located outside of UG, then the correlation that forms the nexus of this paper would be a wholly spurious one. Therefore, the bulk of what follows will be devoted to presenting properties that characterize principles of UG.

The strongest arguments for innate principles of UG are those based on the poverty of the stimulus presented to the child language learner. How could the child have learned such-and-such a principle inductively, one reasons, given its abstractness, the limited amount of relevant information provided, and the speed of acquisition? One's first thought might be that poverty of the stimulus arguments are neutral between identifying UG principles and parsing principles, since both are innate. Nevertheless, the *effect* of a parsing principle should typically be a distinction in, or judgment of, acceptability that *is* learnable through positive evidence alone, while the *effect* of a UG principle should not be. Just as a logical point, there is no reason that distinctions in acceptability whose explanation is based in the structure of the parsing mechanism should present learning difficulties. A learnability problem therefore points directly to an innate UG, not to the parser, which merely employs the existent grammar in language use.

So, for example, Hoekstra and Kooij (1988) motivate Subjacency as a UG principle by pointing out that positive evidence alone could hardly suffice to enable the child language learner to come to the conclusion that (5a) is ambiguous as to the scope of *where*, while (5b) is not:

(5) a. Where did John say that we had to get off the bus?
    b. Where did John ask whether we had to get off the bus?

Interestingly, Freedman & Forster (1985) and Frazier (1985b) have argued on the basis of how Subjacency violations are parsed that this constraint cannot be built into the parser, thereby challenging the earlier claim by Marcus (1980) and Berwick & Weinberg (1984) that it doubles as a UG and a parsing principle.

Poverty of the stimulus arguments also apply (though perhaps not as strongly) to Anaphor Binding. It is by no means evident how positive evidence could lead to the conclusion that (6a) is grammatical, and (6b) ungrammatical, thereby supporting the innateness of Anaphor Binding:

(6) a. John seemed to Mary to help himself.
    b. *John appealed to Mary to help himself.

Now let us turn to EIC. Even though this processing principle, which, again, explains the contrasts of (2) and (3), is innate, there are no poverty of the stimulus arguments applicable to the *acquisition* of these contrasts. Positive evidence can reveal that V-VP-NP, V-AP-NP, and V-PP-NP are options to V-NP-VP, V-NP-AP, and V-NP-PP respectively, as well, I suspect, as the conditions under which the former would be likely to be used instead of the latter.

Or take cases where the effects of EIC have been fully grammaticalized. EIC predicts that more languages are likely to have N-Adj-S' orders than N-S'-Adj, but in this fact I see no evidence for a UG principle, since there is no compelling poverty of the stimulus argument that learners must choose the former order. I assume that if anything is learnable from positive evidence, it is the possible ordering of the constituents of phrases and clauses -- a phenomenon about which EIC has a great deal to say and principles of UG very little.

A second diagnostic for a UG principle is the nature and degree of variation that it exhibits. The current consensus is that each principle admits to variation in a highly circumscribed way, namely by allowing different parameter settings whose values are ordered with respect to one another by the set-theoretic relation of proper inclusion. Take the question of the possible governing categories for Anaphor Binding, where there exists variation from language to language and even between different anaphors within a particular language. Manzini and Wexler (1987) have argued that this variation is highly

systematic: ß can be a governing category for ∂ just in case ß is the minimal category that contains ∂ and a governor for ∂ and has one of the grammatical elements enumerated in (7) (I am oversimplifying somewhat for ease of exposition):

(7) a. a subject; or
    b. an Infl; or
    c. a Tense; or
    d. a 'referential' Tense; or
    e. a 'root' Tense.

Interestingly, the set of categories that have a subject includes the set that has an Infl; the set that has an Infl includes the set that has a Tense; and so on.

While the possible bounding nodes for Subjacency have not been subject to the same cross-linguistic scrutiny as the possible governing categories for Anaphor Binding, the findings of Keenan & Comrie (1977) suggest that very much the same sort of variation might be at work here. They found a universal hierarchy with regard to relative clause formation depicted in (8):

(8) Subject > Direct Object > Indirect Object > Major Oblique Case NP > Genitive NP > Object of Comparison

All languages relativize from subject position; if a language relativizes from a position lower on the hierarchy, it will also relativize from higher positions. It seems probable that their work, translated into current theoretical conceptions, will lead to conclusions about variation in possible bounding nodes that conceptually parallel those of binding domains.

As Manzini and Wexler note, parameterized hierarchically-organized variation is precisely what learnability considerations would lead us to expect. And Piatelli-Palmerini (1989), drawing on the work of J.-P. Changeux, points out that the idea of a UG characterized by a pre-programmed chain of multiple hierarchically ordered internal 'switches' accords well with much current thinking in neurobiology.

Non-UG-related variation in language is quite different. Most importantly, it tends to be graded rather than discrete. Anaphors that are not bound in their governing category in English are impossible, except for those subject to an independent discourse-based condition (Zribi-Hertz 1990), but when is a heavy NP 'too heavy'? To take another example, Dryer (1980) defends at length the generalization that languages are more likely to have sentential noun phrases in final position than in internal position, in initial position than in internal position, and in final position rather than in initial position. Hawkins (to appear) argues that these generalizations are a consequence of EIC. But we do not have here the characteristic property of a UG principle; the choices that languages make in this respect are not matters of choosing the proper 'switch-settings', but rather reflect statistical tendencies which themselves are rooted in the speed with which relevant syntactic nodes can be recognized by the parser.

Surely the graded consequences of parsing principles is a function of the fact that they in turn are rooted to one degree or another in human memory, which itself demands a description along a continuum.

A third characteristic of UG principles is their abstractness. Typically, they manipulate grammatical elements inaccessible to speakers' conscious awareness. So, for example, Subjacency governs the relationship between an overt element and a null element (i.e. a trace), as in overt *Wh*-movement, between two null elements, as in the case of the relationship between an empty operator and its trace, and scopal interpretation where no syntactic movement has occurred at all, as in example (9) from Lakhota, discussed in Foley and Van Valin (1984):

(9)	Wičhaša	wą	takū	ophethu	ki	he	wąlaka	he
	man	a	WH/smthg	3sg-buy	the	that	2sg-see-3sg	Q

They note that this sentence may be interpreted to mean 'Did you see the man who bought something?, but not 'What did you see the man who bought', a fact that follows directly from Subjacency governing (invisible) movement in LF.

Likewise, Anaphor Binding affects null anaphors as well as overt ones. Indeed, the basic insight that led ultimately to the Binding Theory is that the relationship between an anaphor and its antecedent parallels that between the launching and landing sites of NP movements.

I have seen no evidence that non-UG principles ever make reference to null grammatical categories. EIC, for example, treats all NP's the same, whether they are overt or null.

Finally, UG principles tend to be at a further remove from the functional factors that originally motivated them than are functionally-motivated principles not belonging to UG. For example, while Subjacency might have become biologized for the purpose of aiding the hearer to pair antecedents and gaps, there are innumerable sentences that violate this principle (and are therefore ungrammatical) that are nevertheless perfectly easy to understand. Take, for example, a 'Coordinate Structure Constraint' violation such as (10), which in all likelihood can be subsumed under Subjacency:

(10)   *What did John eat beans and?

Not only does this sentence present no processing difficulties, but it forms a minimal pair with grammatical sentence (11), in which Subjacency has not been violated:

(11)   What did John eat beans with?

The divergence over time between selectionally-shaped form and the functional pressures that originally shaped the form is, of course, the norm in evolutionary history. In the general case, form will change to reflect a changing environment only if doing so has a positive effect on the survival and reproductive possibilities of the organism possessing the particular formal structure. Hence, we retain our appendix, even though its utility to digestion has disappeared due to changed diets caused by a changed environment. Fair-skinned people who have migrated to sun-bathed countries show no genetic tendency from generation to generation to darkening of the skin (though if skin cancer rates in Australia are any indication, this might change).

One might speculate then that Subjacency was established as a principle of innate UG before the language faculty supported coordinate structures. Since questioning elements within such structures was never communicatively vital, this principle remained the same. Or alternatively (and perhaps more plausibly), Subjacency might have arisen to disallow certain structures that were difficult to process, yet likely to be uttered, without regard to its effects on parallel -- yet communicatively nonessential -- structures that presented no processing difficulties.

Whichever option is correct, Subjacency illustrates par excellence that central to language design are autonomous grammatical principles that cannot be reduced in toto to the functional principles that in the distant past brought them into being.

There is no question, on the other hand, of parsing principles or principles based, say, on cooperative communication displaying the same sort of form-function disparity. If a sentence were found that was predicted by EIC to create processing difficulties that did not in fact do so, then that would be prima facie evidence that EIC, as a principle of sentence processing, was in need of refinement. Subjacency, as a grammatical principle, is under no such obligation.

In closing, it needs to be stressed that while UG principles point to asymmetry, asymmetry alone does not lead irrevocably to a UG principle. For example, one way to

keep the speaker in check is to build constraints directly into the *production system*. Cutler (1987) and Levelt (1989) discuss innate production constraints whose sole function seems to be to ease communication, the former involving the formation of neologisms and the latter guaranteeing that on-line repairs follow a particular well-formedness rule.

Why then were Subjacency and the other constraints not made production constraints, a solution which, like the UG solution, would effectively prevent sentences like (1) and (2) from being uttered? I suspect that the reason is that the planning units of production (see Levelt) are too short for this to have been a workable possibility.

Finally, asymmetry exerts no pressure toward an innate constraint if it does not lead to communication being impaired. So, as Slobin (1977) has pointed out, speakers and hearers have different 'interests', as far as the packaging of a linguistic form is concerned. The former wish to minimize the articulatory effort in producing it, the latter the effort of understanding. The result is a spectrum of possibilities within a range acceptable to both, from an inflectional 'speaker-oriented' morphology, as in Serbo-Croatian, at one end, to an agglutinative 'hearer-oriented' morphology, as in Turkish, at the other. And in the realm of discourse, Horn (1984) has pointed to a 'division of pragmatic labor' to deal with asymmetries in the needs of speakers and hearers.

Interestingly, some languages have found a way to handle 'awkward' pairings of antecedents and gaps without recourse to a UG principle. This has been accomplished by means of resumptive pronouns, which fill the gap of the displaced *wh*-phrase. Thus in English, Subjacency-violating (and therefore ungrammatical) (12a) becomes grammatical if the gaps are filled with pronouns, as in (12b):

> (12) a.*He's the kind of person who when you meet ___ you like ___.
> b. He's the kind of person who when you meet him you like him.

What the use of resumptive pronouns does is to reduce the degree of asymmetry by making things a little more difficult for the speaker (who has to remember to mark the gap) and quite a bit easier for the hearer (who has the gap marked and can thus link more easily its position to the fronted *wh*-phrase).

Why then, given a possible solution to the asymmetry problem not involving a novel principle of UG, did such a principle arise at all? Why do some languages allow resumptive pronouns, but not others, and why do some languages that have them (like English) restrict their use severely? At the present time, I have no answer to offer to these questions.

To conclude, then, formal autonomous grammatical principles are compatible with a functionalist perspective on language. Indeed, it would seem that the more of a functionalist one is, the more one should be drawn to the idea of autonomous innately-specified universal grammar.

ENDNOTES

*I would like to thank Keith Denning, Matthew Dryer, David Gil, Jack Hawkins, and Willem Levelt for their comments on the subject matter of this paper. None bears any responsibility for the content. For more extended discussion, see Newmeyer (in preparation).

[1]The principle of Subjacency dates from Chomsky (1973) and unifies several of the extraction constraints proposed in Ross (1967). In different ways, Kayne (1984) and Chomsky (1986) attempt to unify Subjacency and the ECP, a result which, if correct, has no bearing on the conclusions of this paper.

REFERENCES

Bates, Elizabeth and Brian MacWhinney. 1979. A functionalist approach to the acquisition of grammar. Developmental pragmatics, ed. by E. Ochs and B. Schieffelin, 167-211. New York: Academic Press.

Berwick, Robert and Amy Weinberg. 1984. The grammatical basis of linguistic performance. Cambridge, MA: MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. A festschrift for Morris Halle, ed. by S. Anderson and P. Kiparsky, 81-162. New York: Holt.

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris.

Chomsky, Noam. 1986. Barriers. Cambridge, MA: MIT Press.

Creider, Chet. 1979. On the explanation of transformations. Syntax and semantics, Volume 12: discourse and syntax, ed. by T. Givón, 3-21. New York: Academic Press.

Cutler, Anne. 1987. Speaking for listening. Language perception and production, ed. by A. Allport, et al., 23-40. New York: Academic Press.

Dryer, Matthew. 1980. The positional tendencies of sentential noun phrases in universal grammar. Canadian Journal of Linguistics 25.225-262.

Foley, William and Robert Van Valin. 1984. Functional syntax and universal grammar. Cambridge: Cambridge University Press.

Freedman, Sandra and Kenneth Forster. 1985. The psychological status of overgenerated sentences. Cognition 19.101-131.

Frazier, Lynn. 1985a. Syntactic complexity. Natural language parsing, ed. by D. Dowty, L. Karttunen, and A. Zwicky, 129-189. Cambridge: Cambridge University Press.

Frazier, Lynn. 1985b. Modularity and the representational hypothesis. NELS 15.131-144.

Givón, Talmy. 1979. On understanding grammar. New York: Academic Press.

Hawkins, John. To appear. A parsing theory of word order universals. Linguistic Inquiry.

Hoekstra, Teun and Jan Kooij. 1988. The innateness hypothesis. Explaining language universals, ed. by J. Hawkins, 31- 55. Oxford: Blackwell.

Horn, Laurence. 1984. Toward a new taxonomy of pragmatic inference: Q- and r-based implicature. Meaning, form, and use in context: Linguistic applications, ed. by D. Schiffrin, 11-41. Washington: Georgetown University Press.

Kayne, Richard. 1984. Connectedness and binary branching. Dordrecht: Foris.

Keenan, Edward and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. Linguistic Inquiry 8.63-99.

Levelt, Willem. 1989. Speaking. Cambridge, MA: MIT Press.

Manzini, Maria-Rita and Kenneth Wexler. 1987. Parameters, binding theory and learnability. Linguistic Inquiry 18.413-444.

Marcus, Mitchell. 1980. A theory of syntactic recognition for natural language. Cambridge, MA: MIT Press.

Newmeyer, Frederick. In preparation. Functional explanation in linguistics and the origins of language. Language and Communication.

Piatelli-Palmerini, Massimo. 1989. Evolution, selection, and cognition. Cognition 31.1-44.

Ross, John. 1967. Constraints on variables in syntax. MIT Ph.D. dissertation.

Slobin, Dan. 1977. Language change in childhood and in history. Language learning and thought, ed. by J. Macnamara, 185-214. New York: Academic Press.

Tomlin, Russell. 1986. Basic word order: Functional principles. London: Croom Helm.

Zribi-Hertz, Anne. 1990. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. Language 65.695-727.

Reduplicated Verbs in Japanese as Grammatical Constructions

Shigeko Okamoto
California State University, Fresno

Introduction[0]

There are a number of syntactic formulas in Japanese which require reduplication of a verb, adjective, or noun. In this paper, I will consider two verbal reduplicative constructions: _Verb-(i)ni Verb_ and _Verbal koto wa Verbal_ (the verbal here may be a verb or an adjective).[1] (Hereafter, I will refer to _Verb-(i)ni Verb_ as the _ni_ construction and to _Verbal koto wa Verbal_ as the _koto wa_ construction.) I will demonstrate that each of these constructions exhibits syntactic, semantic, and pragmatic idiosyncrasies that cannot be accounted for by a modular grammar which separates semantics and pragmatics from syntax. I will further argue that although these constructions cannot be generated by standard modular grammar, they cannot be treated as genuine idioms or compound verbs because of certain properties they bear. I will conclude that these constructions are best described as 'grammatical constructions', as characterized by Fillmore and others (Fillmore, Kay, and O'Connor 1988, Fillmore 1985, Lambrecht 1984).

Semantic Properties

Semantically, both the _ni_ construction and the _koto wa_ construction are non-compositional, hence they cannot be analyzed according to regular semantic interpretation principles. Literally, or according to compositional semantics, _Verb-(i)ni Verb_ means '(in order) to Verb, Verb'. (Here, I interpret the _Verb-ni_ as a purpose phrase, as it is typically used, as in sentence (1).) Thus, for example, _tabe-ni tabeta_ literally means '(in order) to eat, ate'. However, _Verb-(i)ni Verb_, as a whole, means 'Verb to an extreme extent'. Thus, _tabe-ni tabeta_ means 'ate to an extreme extent', as in sentence (2). Sentence (3) is another example of this construction. (TM = topic marker; OM = object marker)

(1) Taroo wa tosyokan e hon   o   <u>kari-ni</u>   it-ta.
     TM library to book OM   borrow to go Past
  'Taroo went to the library to borrow a book.'

(2) Kinoo   wa   susi o   <u>tabe-ni tabe-ta</u>.
   yesterday   TM   sushi OM   eat   eat Past
  'Yesterday (I) ate sushi to an extreme extent.'

(3) Hanako wa   Yosio o   <u>nikum-ini nikun-da</u>.
       TM   OM hate   hate Past
  'Hanako hated Yoshio to an extreme extent.'

As for the expression _Verbal koto wa Verbal,_ on the other hand, compositional semantics yields its literal meaning as 'That . . . Verbal, Verbal'. For example, _oisii koto wa oisii_ literally means 'That (something) is tasty, is tasty', which does not make much sense. However, _Verbal koto wa Verbal,_ as a whole, has its own meaning: Although it is difficult to translate into English, it roughly means 'It is true/I admit that . . . Verbal'. Thus, _oisii koto wa oisii,_ as a whole, means 'It is true/I admit that (something) is tasty', as in sentence (4). Sentence (5) is another example.

(4) Ano   okasi wa <u>oisii   koto   wa   oisii</u>.
    that   cake TM tasty   tasty
   'It is true that that cake is tasty.'

(5) Sakubun   o   <u>kai-ta   koto   wa   kai-ta</u>.
    composition OM   write Past   write-Past
    'It is true that (I) wrote a composition.'

These facts suggest that both *Verb-(i)ni Verb* and *Verbal koto wa Verbal* are like words in that each construction, as a whole, has a specific meaning.

<u>Morpho-syntactic Properties</u>

The two constructions also exhibit morpho-syntactic (and phonological) idiosyncrasies. In the following I will first discuss the properties of the *ni* construction, and then those of the *koto wa* construction.

Some properties of the *ni* construction suggest that it cannot be analyzed as a complex phrasal expression generated by standard phrase structure rules--i.e. as consisting of two independent phrases, the *Verb ni* and the *Verb*, and that it is rather an integral whole like a compound verb. The following five properties will demonstrate this point.

First, the phrase *Verb-(i)ni* is usually interpreted as a purpose expression and followed by a verb of motion, such as *iku* 'go' and *kaeru* 'return', as shown in sentence (1). Standard phrase structure rules do not include a construction in which the phrase *Verb-(i)ni* is followed by the same verb as in the *ni* construction.

Second, as I pointed out above, it is semantically like a single word and resists a componential analysis, hence a regular categorial syntactic analysis.

Third, the *ni* construction does not allow intrusion by an adverbial or a noun phrase, as shown in (6b) and (7b). If we replace the *Verb-(i)ni* in the *ni* construction with a regular manner-adverbial like *omoikkiri* 'as much as one wishes' or *sinsoko kara* 'from the bottom of one's heart', other elements can be placed between the adverbial and the verb, as shown in (6c) and (7c). Thus, if we can regard the *Verb-(i)ni* in the *ni* construction as a regular adverbial phrase, intrusions, as in (6b) and (7b), should be possible.

(6) a. Kinoo   wa Yamada to akegata made nom-ini non-da.
       yesterday TM   with dawn until drink drink-Past
       'Yesterday, (I) drank to an extreme extent with Yamada until dawn.'

    b. *Kinoo   wa   Yamada to nom-ini <u>akegata made</u> non-da.
        yesterday   TM   with drink dawn   until drink Past
       'Yesterday, (I) drank to an extreme extent with Yamada until dawn.'

    c. Kinoo   wa Yamada to   omoikkiri   <u>akegata made</u> non-da.
       yesterday TM   with as much as we wish dawn   until drink Past
       'Yesterday, (I) drank with Yamada, as much as (we) wished, until dawn.'

(7) a. Hanako wa Yosio o   nikum-ini   nikun-da.
              TM   OM hate   hate Past
       'Hanako hated Yoshio to an extreme extent.'

    b. *Hanako wa nikum-ini <u>Yosio o</u>   nikun-da.
              TM hate   OM hate Past
       'Hanako hated Yoshio to an extreme extent.'

    c. Hanako wa sinsoko kara   <u>Yosio o</u> nikun-da.
              TM from the bottom of the heart   OM hate Past
       'Hanako hated Yoshio from the bottom of (her) heart.'

Fourth, the causative form *-(s)ase* applies to the *ni* construction as a whole (e.g. (8a)); it becomes less acceptable if the causative form is applied separately to each instance of the verb (e.g. (8b)).

(8) a. Kooti wa   Taroo o   hasir-ini hasir-<u>ase</u>-ta.
       coach TM   OM   run   run Caus Past
       'The coach made Taroo run to an extreme extent.'

    b. ??Kooti wa   Taroo o   hasir-<u>ase</u>-ni hasir-<u>ase</u>-ta.
       coach TM   OM   run Caus run Caus Past
       'The coach made Taroo run to an extreme extent.'

Fifth, as can be seen in the above examples, tense is marked for the *ni* construction as a whole, and not for each instance of the verb.

The properties listed above may suggest that instances of the *ni* construction be regarded as compound verbs. However, there are also properties that indicate that such a treatment is not appropriate. The following five points will show that the *ni* construction differs from a prototypical compound verb.

First, *Verb-(i)ni Verb* does not have the pitch pattern characteristic of a compound verb: The latter has one accentual peak, as shown in (9a), whereas the former has two accentual peaks--i.e. one peak in each verb, as shown in (9b).

(9) a. ‾‾‾‾
       oti-tuk-u
       fall arrive
       'calm/settle down'

    b. ‾ ‾
       nom-ini nom-u
       drink   drink
       'drink to an extreme extent'

Second, the *ni* construction contains a particle (i.e. *-(i)ni*) which is not found in regular compound verbs. (See examples (9a) and (9b).)

Third, unlike regular compound verbs, the *ni* construction rejects the so-called subject honorific form *o(-Verb)-(i)ni nar(u)*. For example, in (10a), this honorific is appropriately used for the compound verb *kaki-ageru* 'finish writing', whereas in (10b), which contains the *ni* construction, the use of the same honorific is not acceptable.

(10) a. Sensei   wa hon   o   <u>o</u>-kak-i-age -<u>ni nat</u>-ta.
        teacher TM book   OM   write finish HN Past
        'The teacher finished writing the book.'

     b. *Sensei   wa o-sake o   <u>o</u>-nom-ini nom -<u>ini nat</u>-ta.
        teacher   TM   sake OM   drink   drink HN Past
        'The teacher drank sake to an extreme extent.'

Fourth, compound verbs and the *ni* construction behave differently with respect to the passive form *-(r)are*. In the case of compound verbs, the passive form applies to a compound as a whole (e.g. (11a)), but not to each verb in the compound (e.g. (11b)). In the *ni* construction, the passive form can be applied to the construction as a whole (e.g. (12a)). But it can also be applied separately to each instance of the verb (e.g. (12b)).[2]

(11) a. Taroo wa   Hanako ni   tob-i-tuk -are-ta.
         TM   by   jump stick Pass Past
     'Taroo was jumped at by Hanako.'

   b. *Taroo wa Hanako ni tob -are   tuk-are-ta
          TM   by   jump Pass stick Pass Past
     'Taroo was jumped at by Hanako.'

(12) a. Yamada wa   sono kaigi   de   tatak-ini tatak-are-ta.
          TM   that meeting in   attack   attack Pass Past
     'Yamada was attacked/criticized in that meeting to an extreme extent.'

   b. Yamada wa sono kaigi   de   tatak-are-ni tatak-are-ta.
          TM   that meeting in   attack Pass   attack Pass Past
     'Yamada was attacked/criticized in that meeting to an extreme extent.'

Fifth, unlike lexical compound verbs,[3] the *ni* construction is used productively, although there are certain constraints for its application.[4]

I will now discuss the properties of the *koto wa* construction. Like the *ni* construction, the *koto wa* construction has properties which indicate that it is syntactically an integral whole. The following three properties support this point.

First, as I mentioned earlier, *Verbal koto wa Verbal* is semantically like a single word and resists a componential analysis, and hence a regular categorial syntactic analysis.

Second, and related to the first property, the *koto wa* construction cannot be generated by standard phrase structure rules. Normally, the word *koto* is used as a complementizer which nominalizes the preceding clause, and the particle *wa* marks this nominalized clause as the topic. What follows the *koto wa* is the comment. This is not the case in the *koto wa* construction. For example, in (13) and (14), the whole clause preceding the *koto* is the topic and what follows the *koto wa* is the comment; at the same time, in (13), the clause preceding the *koto wa* is the subject of the adjectival expression *tasika da* 'is certain', and in (14), it is the object of the verb *mitomeru* 'admit'. Thus, in sentences, such as (13) and (14), there is a major syntactic (as well as semantic and pragmatic) break after the *koto wa*. (SM = subject marker; CP = complementizer)

(13) Taroo ga ano tegami o   kai-ta   koto   wa   tasika-da.
        SM   that letter   OM write Past CP   TM   certain Pres
     'It is certain that Taroo wrote that letter.'

(14) Ano tegami o   kai-ta   koto wa   mitomer-u.
        that letter OM   write Past CP TM   admit Pres
     '(I) admit that (I) wrote that letter.'

(15) Boku wa   soko e   it-ta   koto wa it-ta.
        I   TM   there to go Past   go Past
     'It is true that I went there.'

In the *koto wa* construction, on the other hand, the 'clause' preceding the *koto wa* is not the topic of the sentence. The topic of the sentence, if any, is the noun phrase marked by the particle *wa*; and the *Verbal koto wa Verbal*, as a whole, functions like a main verbal. For example, in sentence (15), the expression preceding the *koto*, namely, *boku wa soko e itta* 'I went there' is not the topic clause because the *itta* after the *koto wa* cannot be considered the comment of this clause, and also because the *boku* in (15) is followed by the topic marker *wa*.[5] (In (13), on the contrary, the *Taroo* is part of the

clause nominalized by the *koto*, hence it cannot be marked by *wa*.) Thus, in (15), *boku* is the topic as well as the subject, and the *itta koto wa itta* is the main verbal. There is no major break after the *koto wa* in (15), as in the case of (14).

Third, the *koto wa* construction generally rejects intrusion by a manner/time/place adverbial or by a noun phrase, as in sentence (16b). Accordingly, we cannot identify the *Verbal koto wa* with a regular adverbial phrase, such as *hontoo ni* 'indeed', which allows other element to follow it before the final verb, as in (16c). It should be noted, however, that the *koto wa* construction allows intrusion by a few modal adverbials, such as *tasika ni* 'certainly' and *hontoo ni* 'indeed/truly', as shown in sentence (16d).

(16) a. Kinoo   ame ga   fut-ta   koto wa fut-ta.
        yesterday rain SM   fall Past   fall Past
        'It is true that it rained yesterday.'

     b. *Ame ga   fut-ta koto wa <u>kinoo</u>   fut-ta.
        rain SM   fall Past   yesterday   fall Past
        'It is true that it rained yesterday.'

     c. Ame ga   hontoo ni <u>kinoo</u>   fut-ta.
        rain SM indeed   yesterday   fall Past
        'It indeed rained yesterday.'

     d. Kinoo   ame ga   fut-ta koto wa <u>tasika ni</u> fut-ta.
        yesterday rain SM   fall Past   certainly fall Past
        'It is certainly true that it rained yesterday.'

The properties listed above suggest that the *koto wa* construction has word-like properties. However, as in the case of the *ni* construction, the *koto wa* construction cannot be regarded as a compound verb for the following seven reasons.

First, the pitch pattern of the *koto wa* construction differs from that of a compound verb: The former has three accentual peaks, while the latter has only one.

Second, unlike compound verbs, the *koto wa* construction contains the complementizer *koto* and the particle *wa*.

Third, the so-called subject honorific form *o(-Verb)-ni naru* cannot be applied to the *koto wa* construction as a whole (e.g. (17a)), but it must be applied separately to each instance of the verb (e.g. (17b)). The opposite is the case in compound verbs.

(17) a. *Sensei wa tegami o   <u>o</u>-yon-da koto wa yom-<u>ini nat</u>-ta.
        teacher TM letter OM   read Past   read HN Past
        'It is true that the teacher read the letter.'

     b. Sensei wa tegami o   <u>o</u>-yom-<u>ini nat</u>-ta koto wa <u>o</u>-yom-<u>ini nat</u>-ta.
        teacher TM letter OM   read HN Past   read HN Past
        'It is true that the teacher read the letter.'

Fourth, contrary to the case of compound verbs (e.g. (11a, b)), the passive form *-(r)are* must be applied separately to each instance of the verb in the *koto wa* construction, as shown in (18b), and not to the construction as a whole, as in (18a).

(18) a. *Boku wa Taroo ni   nagut-ta koto wa nagur-<u>are</u>-ta.
           I TM   by hit Past   hit Pass Past
        'It is true that I was hit by Taroo.'

b. Boku wa Taroo ni   nagur-<u>are</u>-ta koto wa nagur-<u>are</u>-ta.
   I   TM   by hit Pass Past   hit   Pass Past
'It is true that I was hit by Taroo.'

Fifth, unlike compound verbs (e.g. (19)), in the *koto wa* construction the causative form *-(s)ase* must be applied separately to each instance of the verb, as shown in (20a) and (20b).

(19) Taroo wa   Hanako o   oti-tuk   -<u>ase</u>-ta.
     TM   OM   calm down   Caus Past
'Taroo made Hanako calm down.'

(20) a. *Taroo ni kusuri   o   non-da koto wa nom-<u>ase</u>-ta.
         OM medicine OM   take Past   take Caus Past
'It is true that (I) had Taroo take the medicine.'

    b. Taroo ni   kusuri   o   nom-<u>ase</u>-ta   koto wa nom-<u>ase</u>-ta.
       OM medicine   OM   take Caus Past   take Caus Past
'It is true that (I) had Taroo take the medicine.'

Sixth, unlike compound verbs, each verb in the *koto wa* construction is marked by tense. And, last, the *koto wa* construction is used fully productively.

To summarize the morpho-syntactic (and phonological) properties of the two constructions examined above: Both constructions possess word-like properties as well as complex phrasal properties; neither construction can be regarded as a prototypical compound verb nor a complex phrasal expression generated by standard phrase structure rules. It should be noted, however, that the two constructions are not exactly the same with respect to the degree of their 'word-likeness'. As the analysis above indicates, the *ni* construction is more word-like than the *koto wa* construction.

Semantic and Pragmatic Functions

In what remains, I will discuss the semantic and pragmatic functions of the two constructions. As the examples given above suggest, the *ni* construction involves primarily the propositional content of a sentence whereas the *koto wa* construction concerns both the propositional content and the modality.

With regard to *Verb-(i)ni Verb*, at the level of propositional content, the construction, as a whole, satisfies simultaneously the function of a verb and that of a manner-adverbial: that is, it refers to a certain action/process as well as to the manner in which the action/process is carried out.

The function of the *koto wa* construction, on the other hand, concerns both the proposition and the modality in that the construction, as a whole, refers to a certain action/process/state as well as to the speaker's attitude toward the propositional content. Through the *koto wa* construction the speaker expresses his/her concessive attitude toward the proposition in question. Accordingly, it is typically (but not always) used parenthetically as a concessive preface to the main assertion, which points out a value contrary to the value stated in the preface. For example, in sentence (21), the first clause which contains the *koto wa* construction is the preface to the second clause (i.e. *yoku-nai* 'not good')--the main assertion. Sentence (22B) can be analyzed in the same manner.

(21) Sakubun   o   <u>kai-ta   koto wa kai-ta</u>   ga,   yoku-nai.
     composition OM write Past   write Past but   good Neg
'It is true that (I) wrote a composition, but (it) is not good.'

(22) A: Kono okasi oisii ne.
     this cake tasty
     'This cake is tasty, isn't it?'

    B: Un, <u>oisii koto wa oisii</u> kedo, takai yo.
       yes tasty   tasty but expensive
       'Yes, it's true that (it)'s tasty, but (it)'s expensive.'

In both sentences (21) and (22B), the preface expresses some positive value regarding the subject matter, whereas the main assertion expresses some negative value. This, however, is not always the case; the reverse is also possible. Sentences (23) and (24B) show that the preface containing the *koto wa* construction expresses a negative value, whereas the main assertion conveys a positive value.

(23) Ano okasi wa <u>takai   koto wa takai</u>   kedo, oisii.
     that cake TM expensive   expensive but   tasty
     'It is true that that cake is expensive, but (it)'s tasty.'

(24) A: Kono uti, honto ni semai n da.
       this house really   small
       'This house is really small.'

    B: <u>Semai koto wa semai</u> ga,   totemo kinoo-teki-da ne.
       small   small but   very   functional Pres
       'It is true that (it)'s small, but (it) is very functional.'

In either case, the proposition in the preface is often something that has been (explicitly or implicitly) asserted or questioned in the preceding discourse (e.g. (22) (24)).[6] And the preface containing the *koto wa* construction functions as kind of a concessive echo statement. The speaker, by repeating the proposition, indicates that s/he accepts concessively its validity. But, then the main assertion follows, which reveals that the speaker believes some contrary value to be also applicable.

Compare now sentences (22B) and (25Bi) which differ from each other only in the conjunctions: *kedo* 'but' in (22B) and *kara* 'so/therefore' in (25Bi).

(25) A: Kono okasi oisii ne.
       this cake   tasty
      'This cake is tasty, isn't it?'

    Bi: *Un, <u>oisii koto wa oisii kara</u>, takai yo.
       yes tasty   tasty so   expensive
       'Yes, it's true that (it)'s tasty, so (it)'s expensive.'

    Bii: Un, <u>oisii kara</u> takai n da yo.
        yes tasty so   expensive
        'Yes, (it) is tasty, so (it)'s expensive.'

In (22B), the *koto wa* construction and the conjunction *kedo* are appropriately used: B, the speaker, first agrees with A's opinion concessively, and then asserts his/her own opinion, which points out some value perceived as contrary to the value indicated in A's opinion. In (25B), on the other hand, the use of *kara* indicates that the speaker perceives the first proposition, or A's assertion, as the reason that can justify the second proposition, his own assertion. In other words, A's assertion is treated as compatible with B's own assertion. Accordingly, there is no need for B to concede A's opinion, hence the use of

the *koto wa* construction in (25Bi) is inappropriate. Sentence (25Bii), on the other hand, is acceptable, because the *koto wa* construction is not used.

(26)   A: Kono okasi takai ne.
            this   cake   expensive
            'This cake is expensive, isn't it?'

            B: Un, demo   <u>oisii koto wa oisii kara</u>, takai n da yo.
               yes but   tasty   tasty so   expensive
               'Yes, but it's true that (it)'s tasty, so (it)'s expensive.'

(27)   A: Kono okasi oisii ne.
            this   cake   tasty
            'This cake is tasty, isn't it?'

            Bi: *Un, demo <u>takai koto wa   takai   kara</u>, oisii n da yo.
               yes   but   expensive   expensive so   tasty
               'Yes, but it's true that (it)'s expensive, so (it)'s tasty.'

            Bii: Un,   <u>takai   kara</u>   oisii n da yo.
               yes   expensive so   tasty
               'Yes, (it)'s expensive, so (it)'s tasty.'

Note, however, that there are cases in which the *koto wa* construction can cooccur with the conjunction *kara*.[7] For example, in (26B) the second proposition is the same as A's opinion, which points out a negative value of the subject matter (i.e. the cake). The first proposition in (26B) can be regarded as a positive aspect of the same subject matter which is recognized by people in general. The speaker of (26B) justifies the negative value in question by concessively admitting the positive value in the first proposition as the reason for the negative value. Thus, the use of the *koto wa* construction with the conjunction *kara* is appropriate in (26B).

Compared to (26B), (27Bi) is odd: In (27Bi), the second proposition, which is the same as A's opinion, is a positive aspect of the subject matter, while the first proposition, which is used as the reason for the second proposition, is a negative aspect. There is no need for B, the speaker, to be concessive about the negative aspect in order to justify the positive aspect. Thus, the use of the *koto wa* construction is inappropriate. Sentence (27Bii), on the other hand, is acceptable as a response to (27A) because the *koto wa* construction is not used.

Conclusion

I have pointed out various morpho-syntactic, semantic, and pragmatic idiosyncrasies of the *ni* construction and the *koto wa* construction. These idiosyncrasies make the two constructions idiomatic. Yet, they cannot be treated as genuine idioms because they are used productively: their members are not exhaustively listable in the lexicon. Like many other syntactic formulas (Lambrecht 1984, Fillmore *et al.* 1988), the properties of these two constructions support the view of language that productivity and idiomaticity are not necessarily mutually exclusive. It is also the case that these constructions cannot be fully described by a modular grammar which separates semantics, pragmatics, and lexicon from syntactic rules. I argue that these constructions, along with other reduplicative constructions in Japanese (e.g., *V-temo V-temo, N wa N to site*), are best accounted for by a non-modular construction grammar--i.e. as partially lexicalized integral morpho-syntactic patterns bearing unique semantic and pragmatic properties.

[1]The use of the terms 'verb' and 'verbal' here is at issue. I use the term 'verb' (in the *ni* construction) to refer to the stem form of a verb, and the term 'verbal' (in the *koto wa* construction) to refer to the verb or adjective stem form plus tense.

[2] As the following examples indicate, the acceptability of the passive form -*(r)are* in the *ni* construction seems to differ depending on the verb.

   (i) a. ??Boku wa   sensei ni   sikar-ini sikar-<u>are</u>-ta.
           I   TM   teacher by   scold   scold Pass Past
          'I was scolded by (my) teacher to an extreme extent.'
      b.   Boku wa sensei   ni   sikar-<u>are</u>-ni sikar-<u>are</u>-ta.
           I   TM teacher   by   scold Pass scold Pass Past
          'I was scolded by (my) teacher to an extreme extent.'
  (ii) a.   Boku wa sensei ni   home-ni home-<u>rare</u>-ta.
           I   TM teacher by   praise   praise Pass Past
          'I was praised by (my) teacher to an extreme extent.'
      b. ?Boku wa sensei ni   home-<u>rare</u>-ni home-<u>rare</u>-ta.
           I   TM teacher by praise Pass   praise Pass Past
          'I was praised by (my) teacher to an exteme extent.'

[3]Note that there are so-called syntactically formed, or productive, compound verbs. These compounds are semantically transparent: The second members in these compounds usually express aspectual meanings, and are close to auxiliary verbs (e.g. *tabe-owaru* 'finish eating', *tabe-hazimeru* 'start eating').

[4] For example, the *ni* construction rejects nondurative verbs, such as *tuku* 'arrive' and *tomaru* 'stop' (e.g., (i)); nor does it allow compound verbs containing the verb *suru* 'do' (e.g. (ii)).

   (i) *Boku-tati wa tyoozyoo ni tuk-ini tui-ta.
           we   TM summit   reach   reach Past
      'We reached the summit to an extreme extent.'
  (ii) *Kinoo   wa kaimono-si-ni   kaimono-si-ta.
        yesterday   TM shopping-do   shopping-do-Past
      'Yesterday, (I) did shopping to an extreme extent.'

[5]As will be discussed later in this paper, it is true that the proposition in the 'clause' preceding the *koto wa* is often something that has been recognized in the preceding discourse. Yet, it cannot be considered the topic of the sentence, because the second verb in the *koto wa* construction by itself is not the comment, and also because the whole sentence including the second verb is kind of an echo statement.

[6]I owe this observation to Knud Lambrecht.

[7]This point was brought to my attention by Yoshiko Matsumoto.

References

Fillmore, Charles. 1985. Syntactic intrusions and the notion of grammatical construction. BLS 11. 73-86.
Fillmore, Charles, Paul Kay, and Mary C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. Lg. 64. 501-538.
Lambrecht, Knud. 1984. Formulaicity, frame semantics, and pragmatics in German binominal expressions. Lg. 60. 753-796.

Copula Contraction And Absence In Barbadian English, Samaná English And Vernacular Black English*

John R. Rickford and Renee Blake
Stanford University

This paper is part of a general attempt by our research group at Stanford to reopen some fundamental theoretical and methodological issues in the analysis of the copula in American as well as Caribbean varieties of English, and to bring some new--and needed--data to bear on such issues. Other papers we have written in recent years deal with American vernacular varieties (McElhinny 1988, Rickford 1989, Rickford et al 1988). This paper examines copula contraction and deletion in mesolectal Caribbean English, as represented in casual speech data from six Barbadian speakers.

Although studies of the copula in Caribbean English have been available for some time now, and comparisons with Vernacular Black English (VBE) have become more frequent in recent years, there are no quantitative, accountable descriptions of the Caribbean English copula comparable in sophistication and scope to those available for VBE.

Table 1, for instance, is from Bickerton's (1973) study of the Guyanese copula. To its credit, this study was corpus-based and accountable, in the sense of reporting all variant forms used by each speaker, rather than only those considered representative of the dialect. In these respects, and in its use of the dynamic/implicational framework pioneered by DeCamp (1971) and C.J. Bailey (1973), it was a significant advance on the useful but brief characterizations of the basilectal creole copula which B. Bailey (1965) and Stewart (1969) had provided. However, Bickerton's study was not quantitative.[1] A speaker who used one token of zero and ninety-nine tokens of iz before adjectives would be represented no differently (by a 1/3 entry in column 6, table 1) than a speaker whose distribution was the exact opposite (ninety-nine tokens of zero and one of iz). And while it is helpful to know how often such 1/3 patterns were manifested by different speakers, as against the categorical zero-only (1) or iz-only patterns (3), this is less informative than a full-fledged quantitative study. Another problem was the number of empty cells in Bickerton's study (see Pavone 1980).

Holm's (1976, 1984) study of the copula in Jamaican and Gullah, although serving as a springboard for one of the most insightful discussions of the creole origins of VBE (see also Baugh 1979, Labov 1972), was also limited in several respects. It was based on secondary data from a small number of speakers recorded years earlier (in the case of Jamaica, a single lower mesolectal speaker whose Nansi stories were included in LePage and DeCamp 1960) and it involved some analytical and counting decisions which Holm himself acknowledged (1984:303, n 3) to have been inappropriate:

> After having worked on Miskito Coast Creole ... I realize that table 1 reflects some naive assumptions which I held when this paper was first written. Not all words corresponding to forms of standard English *be* should have been lumped together; [iz] should have been treated separately (as the equative copula before noun phrases), while [wɔz] and [bin] simply mark anterior tense and have little to do with the copula beyond their etymology.

However, no revised analysis of the Jamaican and Gullah data has yet appeared in print. Three Stanford students (Jennifer Knobel, Diana Loo and Michelle Robinson, as part of a class presentation in Linguistics 73, "Black English," 1987) have recoded the data from the source material and provided a preliminary recalculation of the statistics in table 2, but we need to check the recodings and calculations before publishing them. In the meantime we will give one example, in the conclusion of this paper, of how dramatically the reanalysis process can affect Holm's data, which have been accepted as the standard of reference for Caribbean copula absence.

Finally, like virtually all other work on the Caribbean copula to date (Edwards 1980 and Escure 1981 are two other studies which we can only mention in passing here) Holm's study does not examine copula _contraction_, and its copula absence data (see table 2) are for following grammatical environment only. The advantage of a multivariate analysis, of the sort provided by the variable rule program (VARBRUL), is that the effects of other constraints, such as preceding grammatical and phonological environment, can be simultaneously examined (see Rickford 1990).[2]

These limitations in previous studies of the Caribbean copula are significant in their own right, and because they imperil comparisons with VBE. For instance, Poplack and Sankoff (1987) provide a quantitative, variable rule analysis of the English spoken in Samaná (Dominican Republic) which agrees in a number of respects with earlier analyses of American VBE but not with those of the Caribbean. From this, they conclude that Samaná English is closer to VBE and challenges the creole origins hypothesis, since Samaná English speakers are the descendants of African Americans who emigrated from Philadelphia, New York and New Jersey in the 1820's, and their speech is assumed to be a lineal descendant if not equivalent of African American speech in the early 1800's. However, the Caribbean data which provide the basis for comparison are inadequate, as explained above.

In order to provide a more comparable data base for Caribbean English, we made a series of recordings with some native speakers in Barbados a few years ago and now present a VARBRUL analysis of copula contraction and absence in their speech. Our data base is very similar to the one used by Poplack and Sankoff (ibid) for Samaná both in terms of sample size (theirs: 494 full, contracted and deleted copula tokens, ours: 522) and number of speakers (theirs: 8, ours: 6). Barbados is an excellent data source for our purposes because its English vernacular is a mesolectal creole--the kind which is most similar to VBE, especially with regard to copula absence (Bickerton 1973, Escure 1981:2, Holm 1984:303), and the kind which is therefore most profitable for attempts to reconstruct the history of VBE (see Rickford 1974). Anticipating the query some might raise of whether Barbadian English is not atypical of the Caribbean (see Hancock 1980, Cassidy 1980), note that the Barbadian English vernacular (as distinct from the normative Barbadian variety which is commonly cited) shares many creole phonological and lexical features with the English vernaculars of other Caribbean territories. Although some creole grammatical features, such as basilectal habitual a, have not been attested in Barbados in modern times, other grammatical features, such as mesolectal habitual doz, are very common there (see Morrow 1984, Rickford 1989).

In this regard, it is significant that Barbadian--in common with the vernacular English of Guyana, Jamaica, Samaná and Trinidad, but not mainland US VBE--allows zero auxiliary and copula (for convenience we'll refer to both as "copulas") with first-person subjects, as in these two examples from Peter, a Barbadian fish-vendor:

(1) I Ø gon be a Rasta.
(2) I Ø tekkin' off de heads.

This is of course consistent with a creole history in which gon, Verb+ing and adjective predicates (as stative verbs) occur without copulas in underlying structure (see Bickerton 1973, DeBose and Faraclas 1988:476), and in which variations in copula presence on the surface are plausibly treated as due to copula insertion rather than deletion. In any event, the frequency of sentences like (1) and (2) prompted us to follow Poplack and Sankoff (1987) in including all potential occurrences of am, is, and are in our analysis, discriminating between them through a person/number factor group. (There was little person/number non-agreement, except for occasional uses of is with plural and first person subjects.) Like them, and like other students of the copula, we discounted is tokens followed by words beginning with s, clause finals, and other invariant or indeterminate cases. Unlike them, however, we did NOT include reduced or full forms of it's, that's and what's in the count, for the same reason Labov (1969) and earlier scholars had excluded

them--the fact that they occur overwhelmingly as frozen, contracted forms.[3] (It should be noted that 84% of the Samaná tokens of these forms (136/162) are contracted, and that they account for fully one third of the copula tokens (162/492, p. 304) in Poplack and Sankoff's analysis.)

Displayed in table 3 are the variable rule (VARBRUL) probabilities for the contraction and absence/deletion of am, is, and are in Barbados and Samaná. In order to make our analysis comparable to earlier analyses of Samaná by Poplack and Sankoff and of VBE by Labov and Baugh, we will present results for contraction and deletion as computed by Labov's method (see table 4), and we will generally ignore the effects of alternative computational methods (see Rickford et al 1988). However, we have analyzed our data according to each of the formulae in table 4, and at various points in this paper, we will comment on the differences, if any, that they make. (The theoretical motivation for the "Labov contraction" and "Labov deletion" formulae is that contraction is a necessary pre-requisite to deletion; hence surface deletions should be included in the numerator for contraction, and full forms should be excluded for the denominator for deletion). Note too that in the logistic model used in this program, probabilities greater than .5 favor rule application; those less than .5 disfavor rule application; and those just about .5 have little or no effect. Parentheses denote results for factors that were not selected as significant by the regression analysis in the VARBRUL program.

CONTRACTION

Looking first at the contraction results for Barbados (the first data column in table 3), we see a major effect exerted by the nature of the SUBJECT, with personal pronouns (like "he" and "they") strongly favoring, and a full NP subject (e.g. "The man") strongly disfavoring contraction. This effect matches the results reported by Labov and others for VBE and the results reported by Poplack and Sankoff for Samaná (as shown in table 3, although their factor groups don't correspond to ours exactly).[4] One reason for separating the personal pronouns from other pronouns (e.g. "this", "there", and "somebody") is that the personal pronouns now all end in stressed vowels, allowing us to determine whether the strong effect of pronouns simply reflects the separate favoring effect of a preceding vowel that Labov (1972) had found. Even in Labov's data (ibid., 103, table 3.3) the pronoun effect for Labov contraction had exceeded the effect of a preceding Noun Phrase vowel. But the independence of the two constraints is even clearer in our Barbados data, since other pronouns, generally ending in consonants, remain somewhat favorable to contraction (.58), while preceding phonological environment, as an entire factor group, was thrown out as insignificant (note the square brackets around the probabilities for a preceding consonant and vowel in the Barbados contraction column, table 3). In future work, we plan to examine the effects of stress on contraction independently (for instance, in JOE's here vs. the RAdio's here, both NPs, where capitalization indicates stressed syllables) in an attempt to unravel the explanation for the Pro/NP effect which has remained something of a mystery for twenty years.

Turning now to the other factor groups, note that in the PERSON-NUMBER factor group the Barbadian data agree with the Samaná data in showing is most favorable to contraction, and are least favorable, with am in between.[5] The disfavoring effect of are on contraction may be due to the fact that true are contraction is normally blocked after consonants, which is not the case with is. (Compare "John's at home" with "The men're at home"; the copula in the latter, though reduced, never forms a single syllable with the noun.)

Continuing down column 1 of table 3, note that a FOLLOWING PHONOLOGICAL ENVIRONMENT was significant for contraction in Barbados but not in Samaná. The favoring effect of a following vowel in our data may relate to the preferred CV phonotactic environment which a following vowel creates, insofar as the copula consonant remaining after contraction can be reinterpreted as the onset of the following syllable, as illustrated in the following sentences:

(3) Joe zover the hill (CV#CVCV...), vs
(4) Joe'z beside the hill (CVC CVC...)

Returning to table 3, note that the hierarchy of FOLLOWING GRAMMATICAL ENVIRONMENTS agrees with Labov's findings for the NYC Cobras and Jets, and Poplack and Sankoff's findings for Samaná The parallelism is especially striking for Barbados and Samaná, which both show __ gonna significantly ahead of __Verb+ing (.91 versus .55, and .90 versus .48 respectively) while the gap between these two environments in Labov's NYC data (Labov 1972:86-87) is much smaller (3 percentage points for the Jets, 4 percentage points for the Thunderbirds). One interesting point about this hierarchy, however, is that it is completely reversed when the data are computed by the "straight contraction" rather than the "Labov contraction" formula: A following NP becomes most favorable, and a following gonna least. This is because the "Labov contraction" hierarchy for following environment derives primarily from the high proportion of copula deletions or absences in the data; once these are removed, the hierarchy collapses. To our mind, this is as it should be; there are valid reasons for following grammatical environments to pattern as they do with respect to copula absence, in terms of prior creole grammatical categories (see Holm 1984:298); but no explanation has yet been proposed for their having a similar effect effect on copula contraction.

If we now turn our attention to the SPEAKER factor group at the bottom of table 3, we see that this factor group was thrown out as insignificant for contraction in the Samaná data. However, Poplack and Sankoff did find a significant speaker effect for copula deletion (see the far right column), leading them to observe (p. 308) that "as expected, it is the process of deletion which has social significance in the community, in contrast with contraction."

However, if we look at the Barbados data, we see that the expectation that contraction would have no external or inter-speaker significance is not sustained. The speaker factor group was significant both for contraction AND deletion, with individual speaker values varying quite significantly in each case. Furthermore, significant effects are obtained for this factor group regardless of the contraction formula used. Now, can we say anything more about this external factor group beyond the fact that individuals vary? There isn't any obvious effect of gender (Mary, the only woman, has VARBRUL results which are almost identical to Peter's), race (Daniel, the only white speaker, is comparable to Mac, a black speaker), nor social class and age (most of our speakers are in their twenties and thirties and clearly working class). The single biggest effect seems, in fact, to be the stylistic level each speaker adopted in the interview, itself a function of various contextual factors, including his or her relation with the interviewer and other interlocutors (Bell 1984). This can be illustrated most dramatically with the example of Cricketman, the captain of a local cricket team, who was fortuitously interviewed while watching a cricket match. Cricketman varied so dramatically in his speech to the interviewer versus his speech to his teammates that we've identified him as Cricketman 1 and Cricketman 2 in the transcript which follows, and also in the analysis (see table 3). He really behaved like two different people in each persona:

(5) From an interview in August 1987. (Cricketman 1= dialogue with interviewer, Renee Blake; Cricketman 2=dialogue with peers; countable instances of copula in his speech are underlined.)

Cricketman 1: Because right now you are in de northern part of de island--St. James, St. Peters, St. Vincent--St. Andrews, St. Thomas. Now some of dese guys are from de southern side which would be St. Michael--the batting side--most guys will be, Christ's Church. You play on a sort of--you know, but de guys are interzonal. Ya understand?
Interviewer: Right, right. You're from where? St. James?
Cricketman 1: Yeah, yeah. . . .

impressionistic claims for Caribbean varieties made by earlier researchers. Note too, that these results are robust, unaffected by the computational formula used.

The deletion results for the FOLLOWING GRAMMATICAL factor group are also robust. Here the hierarchy is clearly <u>Gonna</u>, Verb+<u>ing</u>, Loc., Adj., and NP (way behind) regardless of the method we use. The inexplicable, only slightly disfavoring effect of NP which Poplack and Sankoff had found is not replicated in our data, but our Locatives are ahead of Adjectives, as in their data, and in Labov's Jets data. Interestingly enough, the 'high Adj.' (over Loc) pattern which Baugh found for deletion in Labov's data for the Cobras and in his own Los Angeles data was not replicated.

For the sake of comparability, Figure 1 displays relative **frequencies** (not probabilities) for copula absence (computed as "**straight deletion**") in the Barbados and Samaná data sets, and in the NYC Jets and Jamaican data analyzed by Labov (1972:86) and Holm (1984:86) respectively. The NYC Jets and Barbados patterns are parallel throughout the range, except that the Barbados data show a bigger NP vs. Adj effect, comparable to that which obtains in the Jamaican data. (Adjectives, of course, are really a subcategory of verb in creole grammar, and require no copula, while noun phrases are quintessential statives and almost always require a copula, whether creole or English-derived.) The Samaná data resemble both the Barbados and NYC Jets data in the relative ordering of the various environments, although the absolute frequencies are lower.

Beyond the initial similarity of their NP vs Adj effect, Jamaican and Barbados diverge sharply, but we have reason to suspect that this divergence is more apparent then real, a function of the fact that Holm's Jamaican figures include percentages for basilectal creole markers (like preverbal <u>de</u>) which are excluded from serious alternation with $\emptyset$ and the inflected copula once a certain level of the continuum is reached. In the Verb+<u>ing</u> case, for instance, only $\emptyset$ and inflected <u>is</u> or <u>are</u> can occur in equivalent syntactic slots; basilectal <u>de</u> and <u>a</u> cannot co-occur with Verb+<u>ing</u> (*"dem <u>de</u> waakin"), but only with Verb ("dem <u>de</u> go"), and therefore tokens with these variants should not be considered along with the others. However, 82% of the variants in Holm's preverbal subcategory for Jamaican (see table 2) come from <u>de</u> and <u>a</u>; if these are removed, leaving only tokens of inflected <u>be</u> and $\emptyset$, the proportion of zero for Verb+<u>ing</u> climbs to 89%, matching the Barbados data. A similar categorization or computation error probably account for the low <u>gonna</u> figure which Holm (ibid.) reports for Jamaica (32%, p. 293), possibly a failure to separate <u>gonna</u> (derived from <u>goin' to</u> and therefore capable of showing variation with $\emptyset$ and <u>be</u>) from <u>go</u>. As Holm himself observes (ibid., p.298), <u>go</u> is a preverbal irrealis or future tense marker which was never preceded by any copula-like particle in the creole. (Even in VBE, we have found, following a suggestion of Raina Jackson's, that <u>gon</u> as in "He gon tell," shows a higher proportion of copula absence than <u>gonna</u>, as in "He's gonna tell.") Although we haven't completed all the necessary reanalyses and recalculations of DeCamp's Jamaican data originally examined by Holm, we expect them to show copula absence figures for <u>gon</u> as high as in our Barbados data and in line with the other data sets. In short, we expect the parallelism between the four data sets of figure 1 to be even stronger.

SUMMARY AND CONCLUSION

In this paper, drawing on recently collected Barbadian data, we have provided the first data-based discussion of copula contraction in Caribbean English, and the first quantitative, VARBRUL analysis of copula contraction and absence in a Caribbean English variety other than Samaná. The results are interesting in their own right, but are especially significant for the challenges they offer to the conclusions of Poplack and Sankoff (1987) that copula contraction and absence in Samaná and VBE are similar, and different from the creole or Caribbean patterns established by Holm (1984).

For contraction, as computed by Labov's formula (see table 4), we note a number of striking parallels between Barbadian, Samaná and VBE, especially with regard to the effect of the grammatical subject, the copula form in question (<u>am</u> vs <u>is</u> vs <u>are</u>), and the following grammatical environment. Thus, the contraction similarities which Poplack and

Cricketman 1: Yeah, yeah. . . .
Cricketman 2: Hit de ball through de fielders, man! Marpuh, wha yuh Ø doin'?!
Interviewer: Who's winning now?
Cricketman 1: Well--i'--i'--i' i' still--de game is not at a stage of winning (?). It's a question of--de battin' side has a hundred and something runs to knock off, so it's not--de game is not at a stage.
Interviewer: Oh. I see what you're saying.
Cricketman 2: Go on Marpuh! Daz a straight ball! Daz a straight ball! Man, he Ø out! Daz a straight ball! Dat ball ain't turn nowhere! He Ø out!

Of the eight potential copula tokens which occur in this extract (there are also five tokens of contracted it's and daz which were treated as "Don't Count" [DC] forms for the reasons given earlier), five occur in the speech of Cricketman speaking to the interviewer-- and they are all full forms (is, are), while three occur in the context of the animated Cricketman yelling onto the field--and these are all zeroes or deletions, contributing to the contraction count according to Labov's formula (table 4). Overall, Cricketman's contraction probabilities vary from .22 to .80 in the two modes, and his deletion values from .14 to .88. (Comparable differences remain even with other contraction and deletion formulae.) Although Poplack and Sankoff were sensitive to possible interviewer and stylistic effects and carried out "standard sociolinguistic checks for detecting more vernacular styles," they may not have fully estimated the potential size of the interlocutor effect.[6] In general, the high and low figures for contraction and deletion in our data seem to correlate most clearly with the degree of rapport which was established between the interviewer and interviewee. After Cricketman 2, the two Barbadians with the highest contraction and deletion probabilities were Peter and Mary. Peter was interviewed by John, a fellow West Indian and Creole speaker, with whom he was able to converse as an insider on numerous Caribbean topics, and Mary was the interviewee with whom Renee had established the closest "homegirl" relationship beforehand, enabling her to discuss various topics at a level of intimacy that strangers would not. By contrast, Mac, who was more formal throughout this interview with Renee, had much lower contraction and deletion rates.

ABSENCE/DELETION

Although the contraction results for Barbados converge with previous work on VBE and Samaná more than the existing literature might have led us to expect, our results for copula absence or deletion diverge from previous work in several respects.

In the SUBJECT factor group (see the Deletion column for Barbados in table 3), the favoring effect which personal pronouns had shown for contraction is completely reversed, with NP now significantly favoring (.84) and personal pronouns disfavoring (.19). This is clearly different from what Labov had found for VBE, and prevents us from arguing for Barbados what Poplack and Sankoff (ibid., 299) do for Samaná: that "deletion is an extension and generalization of contraction." Note, however, that the Samaná data also show a strong NP_ favoring effect. It may be that whatever phonological, stress-favoring effect personal pronouns provide for contraction is lost at the level of deletion; and perhaps that copula absence in these two communities, if not in the Caribbean more generally, is not phonologically constrained at all, but rather should be conceived of as **grammatically conditioned copula insertion** (as many creolists have insisted all along) followed by phonologically conditioned contraction. However, the strong NP_ favoring effect in the Barbadian data disappears when you use "straight deletion" instead of "Labov deletion" (Other Pro becomes .59, and Personal Pronoun and NP .46 and .45 respectively), suggesting, overall, that there is no stable effect of this factor group on deletion--that it may be essentially irrelevant.

The deletion statistics for PERSON, PRECEDING PHONOLOGICAL ENVIRONMENT, AND FOLLOWING PHONOLOGICAL ENVIRONMENT fail to show any significant effects for the Barbados data. This is contrary to the Samaná results for person and preceding phonological environment, but in accord with the impressionistic

263

Sankoff (ibid) demonstrate between Samaná and VBE do not simply line them up on one side, typologically and diachronically, with Caribbean creole vernaculars on the other. Now that contraction data on one Caribbean variety are available, they turn out to be similar to both of these better-documented varieties, providing no opposition to the creole origins hypothesis for VBE.

With respect to copula absence or deletion, Barbadian and Samaná English, in common with other Caribbean vernaculars, both allow for the possibility of zero where standard English and VBE would require full or contracted am (the contracted form is virtually categorical in VBE). In short, Samaná English copula absence is not as different from that of the creole vernaculars as we might otherwise have believed it to be. At the same time, Barbados, Samaná and VBE are all fairly similar in terms of the relative effect of following grammatical environment on copula absence, challenging the stereotypes which have developed with respect to this issue: for instance, that Caribbean vernaculars show high rates of copula absence before adjectives but not Verb+ing or gonna (the so-called "high adj" pattern). As we have shown by reanalyzing Holm's copula absence figures for Jamaica, this stereotype may derive from inappropriate counting decisions which make a following Verb+ing or gonna seem much less favorable to copula absence than they really are.

Clearly, additional quantitative and VARBRUL analyses of copula variation in Caribbean varieties--drawing on original data sets from Jamaican, Guyanese and other territories--remain to be done; we have started on the process, along with others (like Winford 1989). This paper should help to indicate the potential value of such analyses, both for understanding Caribbean vernaculars themselves and for pursuing the decades-old controversy about the roots of American VBE.

FOOTNOTES

*An earlier version of this paper was presented in 1988 at the 17th Annual Conference on New Ways of Analyzing Variation in Language (NWAV 17) held at the University of Montreal at Quebec.

1. This is contrary to the impression which the quotation from Bickerton (1971:491) in Holm (1984:303) might suggest: "In the mesolect, 'deleted' copula is found oftenest with gon (-Fut), not quite so often with -ing forms (Vb) ..."
2. Winford's (1989) conference paper--which we were fortunate enough to see just before finalizing this paper--provides a new analysis of the Trinidadian copula which is data-based, accountable and quantitative. Although it does not employ the variable rule program or any similar multivariate procedure to estimate the individual effects of proposed constraints, its univariate analyses of the effect of preceding and following grammatical environments are extensive and interesting, as are its comparisons with Labov's VBE data.
3. Reduced forms of dat's occur most often as daz, not das, but some dat's and da also occur.
4. Poplack and Sankoff's study includes only a SUBJECT factor group, with CONTRACTION results for individual factors as follows: NP .08; I .45; he, she .93; it, that, what .85; here, there, where .74; we, you, they .32; those, them, these, this .13. For DELETION, their corresponding results are: NP .81; I .06; he, she .28; it, that, what .06; here, there, where .53; we, you, they .90; those, them, these, this .43. These figures should clarify the correspondences between our results.
5. Poplack and Sankoff's data show a bigger are/is difference than ours do, but their is-probability figure is for he and she subjects only, while ours includes these subjects as well as singular NPs.
6. The relatively low frequency of copula deletion which they record for Samaná overall (26%) may not be fully representative of vernacular Samaná (compare 19% for Cricketman 1, to Renee). In any case, Cricketman's data remind us of the need for repeated recordings with varying interlocutors (see Labov 1972, Rickford 1987 and Winford 1972).

REFERENCES

Bailey, Beryl Loftman. 1965. Toward a new perspective in Negro English dialectology. American Speech 40(3): 171-77.

Bailey, Charles-James. 1973. Variation and linguistic theory. Washington, D.C.: Center for Applied Linguistics.

Baugh, John. 1979. Linguistic style-shifting in Black English. Ph.D. dissertation, University of Pennsylvania.

_____. 1980. A re-examination of the Black English copula. Locating language in time and space, ed. by William Labov, 83-106. New York: Academic Press.

Bell, Allan. 1984. Language style as audience design. Language 13: 145-204.

Bickerton, Derek. 1971. Inherent variability and variable rules. Foundations of Language 7: 457-92.

_____. 1973. On the nature of a creole continuum. Language 49: 640-69.

Cassidy, Frederic G. 1980. The place of Gullah. American Speech 55:5-15.

DeBose, Charles E., and Nicholas Faraclas. 1988. An Africanist approach to the linguistic study of Black English: Getting to the African roots of the tense/aspect/ modality and copula systems in Afro-American. Paper presented at the International Round Table on Africanisms in Afro-American Language Varieties, University of Georgia, Athens.

DeCamp, David. 1971. Toward a generative analysis of a post-creole continuum. Pidginization and creolization of languages, ed. by Dell Hymes, 349-70. Cambridge: Cambridge University Press.

Edwards, Walter F. 1980. Varieties of English in Guyana: Some comparisons with BEV. Linguistics 18: 289-309.

Escure, Genevieve. 1981. Decreolization in a creole continuum: Belize. Historicity and Variation in Creole Studies, ed. by Arnold Highfield and Albert Valdman, 27-39. Ann Arbor: Karoma.

Hancock, Ian. 1980. Gullah and Barbadian: Origins and relationships. American Speech 55:17-55.

Holm, John. 1976. Copula variability on the Afro-American continuum. Conference preprints, first annual meeting of the Society for Caribbean Linguistics Turkeyen, Guyana. Compiled by George Cave. Linguistics Section, Department of English, University of Guyana.

_____. 1984. Variability of the copula in Black English and its creole kin. American Speech 59: 291-306.

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. Language 45: 715-62.

_____. 1972. Language in the inner city: Studies in the Black English Vernacular. Philadelphia: University of Pennsylvania Press.

LePage, Robert B., and David DeCamp. 1960. Jamaican Creole. Creole Language Studies, vol 1. London: MacMillan.

McElhinny, Bonnie. 1988. Contraction of auxiliaries in White English. Paper presented at the 17th Annual Conference on New Ways of Analyzing Variation (NWAV17), held at the University of Montreal at Quebec.

Morrow, Todd E. 1984. Bajan. B.A. Linguistics honors thesis, Stanford University.

Pavone, James. 1980. Implicational scales and English dialectology. Ph.D. dissertation, Indiana University.

Poplack, Shana and David Sankoff. 1987. The Philadelphia story in the Spanish Caribbean. American Speech 62: 291-314.

Rickford, John R. 1974. The insights of the mesolect. Pidgins and creoles: Current trends and prospects, ed. by David DeCamp and Ian Hancock, 92-117 Washington, DC: Georgetown University Press.

Rickford, John R. 1987. The haves and have nots: Sociolinguistic surveys and the assessment of speaker competence. Language 16: 149-178.

_____. 1989. Continuity and innovation in the development of BEV be2. Paper presented at the 18th annual conference on New Ways of Analyzing Variation (NWAV-18), Duke University, Durham, North Carolina.

_____. 1990. Response [to Mufwene's Number Delimitation in Gullah, 1986]. American Speech 65.

_____. In press. Grammatical variation and divergence in Vernacular Black English. Internal and External Factors in Syntactic Change, ed. by Marinel Gerritsen and Dieter Stein. (From a 1989 Int'l Conference on Historical Linguistics workshop.) Amsterdam: John Benjamins.

Rickford, John R., Arnetha Ball, Renee Blake, Raina Jackson, and Nomi Martin. 1988. Rappin' on the copula coffin: Theoretical and methodological issues in the variable analysis of BEV contracted and deleted be. Paper presented at the 17th Annual Conference on New Ways of Analyzing Variation (NWAV-17), University of Montreal, Quebec.

Stewart, William A. 1969. Historical and structural bases for the recognition of Negro dialect. Report of the twentieth round table meeting on linguistics and language studies, ed. by James Alatis, 239-48. Washington D.C.: Georgetown University Press.

Winford, Donald. 1972. A sociolinguistic description of two communities in Trinidad. Ph.D. dissertation, University of York.

_____. 1989. Another look at the copula in BE and Caribbean creoles. Paper presented at the 18th Annual Conference on New Ways of Analyzing Variation (NWAV 18), Duke University, Durham, North Carolina.

TABLE 1

TABLE 1. Implicational table for copula distribution (Bushlot).
Key: 1 = *de/bin de* in Cols. 1–4; *a/bina* in Cols. 5, 7, 8; *ø/bin* in Col. 6.
  2 = *ø* except in Col. 6.
  3 = *iz/woz* (no person concord).
  4 = *be* with full person concord.
Environments: Col. 1 = locative; Col. 2 = existential; Col. 3 = time/manner adverbials; Col. 4 = preceding non-finite structures; Col. 5 = cleft S's; Col. 6 = pred. adj.; Col. 7 = NP complement; Col. 8 = impersonal S's; Col. 9 = V-*ing*. Scalability = 95.6%.

SPEAKER	1	2	3	4	5	6	7	8	9
23.	1						1		
16.	1			1				1	
7.	1			1					
20.		1	1						
24.		1							
26.	1	1			1	1			
2.	1	1				1			
9.	1	1			1	1	1	1	
25.		1							
4.						1			
12.						1			
14.						1			
6.						1			
21.						1			
10.						1	1		
28.	1		1	1	1				
3.									2
5.	1			1			1	23	
27.	1		1	1	13	3	3		
15.	1		1	1	13				
17.			1		3	3	3	①️	
1.						3			3
13.	1			13				①️3	
11.	1			3		3	①️3		
18.							3		
19.				4	4				4

Source: Bickerton 1973: table 2, p. 651.

TABLE 2

Ranking of Copulas Favored by
Following Syntactic Environment

A. Jamaican

___V		___Adj		___Loc		___NP	
de	76%	ø	66%	de	45%	be	47%
ø	17%	be	23%	ø	17%	a	31%
a	6%	a	9%	be	17%	ø	22%
be	2%	de	2%	ben	17%	de	0%

B. Gullah

___V		___Adj		___Loc		___NP	
də	46%	ø	62%	bɪn	70%	ɪz	64%
ø	28%	ɪz	13%	ø	22%	bɪn	12%
bɪn	11%	bɪn	13%	də	3%	ø	11%
ə	7%	də	7%	ɪz	3%	wɔz	10%
ɪz	7%	wɔz	7%	wɔz	3%	də	5%

Source: Holm 1984: table 1, p. 292

TABLE 3

Varbrul Probabilities for Labov Contraction and Deletion
of <u>am</u>, <u>is</u>, and <u>are</u> in Barbados and Samaná

FACTOR GROUP	Factors/Constraints	CONTRACTION Barbados	Samaná	DELETION Barbados	Samaná
SUBJECT	Pers. Pro:	.79		.19	
	Other Pro:	.58		.45	
	NP___:	.16	SUBJECT { .08	.84	SUBJECT { .90
PERSON	1st Sg: 'am'	.56	.45	[.47]	.06
NO.	Pl & 2nd Sg: 'are'	.35	.32	[.58]	.53
	3rd Sg: 'is'	.60	.93	[.45]	.28
PREC.	Cons___:	[.48]	.38	[.58]	Not Sig.
PHON.	Vowel___:	[.52]	.62	[.42]	"
FOLL.	___Cons:	.41	Not Sig.	[.54]	.64
PHON.	___Vowel:	.59	"	[.46]	.36
FOLL.	___Gonna:	.91	.90	.77	.59
GRAMM.	___Ving:	.55	.48	.65	.46
	___Loc:	.54	.40	.54	.23
	___Adj:	.40	.35	.42	.19
	___NP:	.16	.24	.08	.41
SPEAKER	Mary:	.64	Not Sig.	.62	.37 #1
	Peter:	.63	"	.65	.63 #2
	Cricketman 1:	.22	"	.14	.79 #3&4
	Cricketman 2:	.80	"	.88	.14 #5
	Mac:	.43	"	.18	.73 #6
	Sarge:	.31	"	.70	.39 #7&8
	Daniel:	.46	"	.34	
Overall %s:		74%	77%	61%	26%
# Of Copula Tokens (n's):		522	489	385	376
Input Probabilities:		.86	.89	.88	.54

Notes:
•Parentheses [ ] indicate values for factors thrownout as insignificant during VARBRUL regression step-up/step-down analysis.
•Personal pronouns: <u>you</u>, <u>she</u>, <u>we</u>, <u>they</u>.   Other pronouns: <u>these</u>, <u>somebody</u>, and so on.
•Cricketman 1: Cricketman to interviewer (RB); Cricketman 2: Cricketman to peers during game.
•Samaná source: Poplack and Sankoff 1987: 306, table 5.
•Variants of <u>it's</u>, <u>that's</u>, and <u>what's</u> included in Samaná data, but not Barbados data.

TABLE 4

Alternative formulae for computing contraction and deletion %'s

Hypothetical data set: 10 tokens of is or are (Full Forms, F), 10 tokens of 's or 're (Contractions, C), 10 tokens of Ø (Deletions, D).

Straight contraction:	$\dfrac{C}{F + C + D}$	$= \dfrac{10}{30}$	$= 33\%$
Straight deletion:	$\dfrac{D}{F + C + D}$	$= \dfrac{10}{30}$	$= 33\%$
Labov contraction:	$\dfrac{C + D}{F + C + D}$	$= \dfrac{20}{30}$	$= 66\%$
Labov deletion:	$\dfrac{D}{C + D}$	$= \dfrac{10}{20}$	$= 50\%$
Romaine contraction:	$\dfrac{C}{F + C}$	$= \dfrac{10}{20}$	$= 50\%$

Source: Rickford et al. 1988: table 1.

FIGURE 1

Copula absence by following grammatical environment (straight deletion freqs.)

FOLLOWING GRAMMATICAL ENVIRONMENT
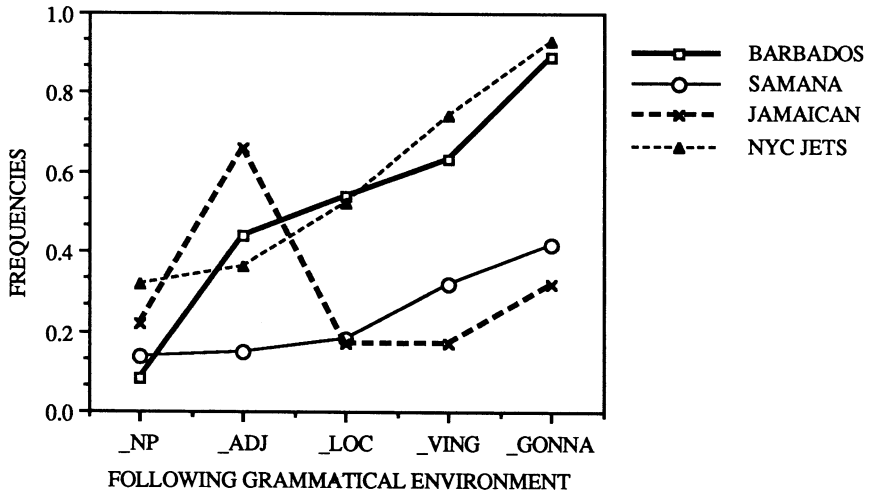
Sources: Samaná, Poplack and Sankoff 1987:305; Jamaica, Holm 1984: 293; NYC Jets, Labov 1972:86.

Parts of Speech in Autolexical Syntax
Jerrold M. Sadock
University of Chicago

Aristotle (De Int 2:20, e.g. Ackrill, 1987) told us that nouns are words that do not express time. Bloch and Trager (1942) said that nouns are words that are centers (i.e. heads) of substantive phrases that may be preceded by modifiers. Langacker (1987) told us that nouns indicate regions in a domain, and Miss Reardon told me that nouns are names of persons places, or things.

None of these claims is wrong, per se, but the problem that I see with each of them is that it is one sided. Aristotle's rule for recognizing nouns is (basically) morphological, since what he meant was that Greek nouns do not take tense inflections. Bloch and Trager provided a completely syntactic criterion (though elsewhere, to be sure, they included morphological touchstones.) Langacker's characterization is semantic, employing entirely cognitive constructs, and my fifth grade teacher's definition was pragmatic, noting the use to which nouns are typically put. It seems clear to me that nouniness involves all of these things, and that similarly, for other parts of speech, characteristics relating to various linguistic dimensions figure in their classification.

The multi-modular view of parts of speech is familiar (see, for example, Schachter 1985), and I will therefore not attempt to back it up, except by example. What I wish to do here is to show that the multi-modular definitions of parts of speech can be given natural and enlightening formulations in a theory, such as Autolexical Syntax, that radically separates the representation of linguistic expressions in the various components of the grammar.

Suppose we take a grammar to be a set modules or components, each of which is itself a grammar of an independent level of linguistic representation (i.e. the "tactics" of that level in the terminology of Stratificational Grammar (Lamb 1966, Lockwood 1972)). The number and nature of the modules needed for the accurate description of natural languages is a complex, partly empirical, and partly theoretical issue, but to begin with, let us assume the existence of three traditional modules: syntax, semantics, and morphology. The syntax specifies the phrasal constituent structures that the language allows, the semantics gives us the set of well-formed meaning structures in the language, and the morphology the set of well-formed morphological entities, less formally: words. I take it to be a virtue of this system of grammar that there is only one autonomous set of semantic principles, and one autonomous set of morphological principles, a virtue absent in many hierarchical theories where both semantics and morphology are split into two or more quite separate components.

Finally let us suppose that, unlike what is assumed in Stratificational Grammar or Transformational Grammar, these modules are not hierarchically related to one another. Conceived of as a grammar of a certain dimension of representation, a module need not wait for the output of another to do its work, but has the power to generate (or analyze) an infinite set of representations quite independently of what is going on in any of the other components. Each component is a self-contained system, with its own independent set of rules, principles, and basic vocabulary.

The glue that binds these independent grammars together and makes them a description of a single language is the lexicon, an annotated list of the fixed expressions in the language, be they morphemes, words, or phrases. Each item on this list, i.e. each lexeme, also includes statements as to its behavior in each of the parallel modules, indicating, for example, whether the item is a morphological stem or affix, whether it combines syntactically with NP complements, and whether it is a predicate or operator in the semantics, and so on. Besides the lexicon as a link between essentially autonomous

representations, a system of general matching constraints comprising an interface module is needed. This interface component sees to it that the topological properties of different dimensions of representations are sufficiently similar to be representations of the same expression, preventing, for example, the association of the syntactic structure of <u>John thinks that Larry left</u> with the semantic structure of <u>Larry thinks that John left</u>.

I will call the basic idea that the topological properties of the autonomous representations of a single expression must be similar the General Homomorphism Constraint. There are two parts to this, one concerning hierarchical structure and the other concerning linear order, which may be informally stated as follows:

(1) General Homomorphism Constraint

a) Constructional Integrity Constraint

Elements associated to two different autonomous representations must bear the same hierarchical relations in both.

b) Linearity Constraint

Elements associated to two different autonomous representations must occur in the same linear order in both.

There are various ways in which these constraints are sometimes relaxed, allowing for the existence of partly non-congruent pairs of representations. I take up the question of constraining mismatches in considerable detail in my forthcoming book (Sadock forthcoming), but for the most part it need not concern us here.

Given some rudimentary modules, such as the ones contained in the appendix to this paper, we will immediately notice that the lexicon will contain not only large classes of lexemes with particular properties in individual modules, but also with similar alignments of properties across the various components. It is the case that the distribution of nouns, for example, is largely consistent within the syntax. To a great extent it is exactly the same items that can take prepositional complements (as per SF9) as refuse NP complements or V-Bar complements, and this is almost exactly the same set that can take preceding determiners according to SF2. To capture such intermodular generalizations, we use the same symbol "N" in various rules such as SF2 and SF9, but a different symbol, "V" in SF4 or SF6.

When we examine the morphology we also find that the lexicon is partitioned according to the distributional privileges of its members. Not all stems can be inflected and derived in all possible ways in a language like English (as opposed to Nootka (Sapir 1921, Swadesh 1939)), so morphological stems will have to be divided into various classes according to the range of morphological processes they can undergo. <u>Reader</u>, and <u>teflon-like</u> are possible words of English, but <u>read-like</u> and <u>tefloner</u>[1] are not. We can extract a long list of items that admit the agentive suffix -<u>er</u>, and reject the adjectival suffix -<u>like</u>, and a long list that have exactly the opposite behavior. True, many phonological stems allow both kinds of suffixes, but these are always ambiguous: There are two very different stems <u>bear</u> in <u>bearer</u> and <u>bear-like</u>, to take an extreme example. We might provisionally call the class that can be suffixed with -<u>er</u> but not with -<u>like</u> A-stems, and the other class B-stems, for we have no grounds as yet for calling them verb stems and noun stems.

But just as soon as we turn our attention back to the syntax, we are struck by the fact that by and large, the morphological A-stems are the Vs in the syntax, and the

morphological B-stems are the Ns. As reasonable and mundane as this association might seem to be, it is not a logical necessity. One could certainly conceive of a language in which the classes obtained on the basis of morphological behavior do not correspond in any way to the classes obtained by noting how things behave in the syntax. There is, after all, Nootka, which Sapir 1921 and Swadesh 1939 described as having only one morphological class up to the inflected word, but at least two classes in the syntax[2]. We could also imagine a language in which all morphological classes are phonologically based, but no syntactic classes are.

But most languages do not work that way, and we are quite secure in the expectation that the major classes of stems in the morphology will turn out to correspond to a large extent to the major classes in the syntax, and thus deserve to be called by the same names that we use for syntactic classes, namely verbs and nouns. We might even elevate this expectation to a principle (and I believe we should), thereby making an important empirical claim:

(2) In the default case, a morphological X will be a syntactic X.

Nevertheless, we must bear in mind that words like "noun" and "verb" are importantly ambiguous, and for that reason, I will regularly distinguish the nouns extracted on the basis of syntactic criteria from those that we get by taking note of morphological facts by using positive bar level symbols for the former, and negative ones for the latter, even for uncomplemented (zero bar) categories.

Thus the lexicon will contain derivational affixes such as (3) and (4) that respectively apply to noun stems (our erstwhile B-stems), or yield noun stems. Therefore teflon-liker is ungrammatical (in the intended sense where telfon-like is a derived adjective), but reader-like, is possible, though not current. The lexicon will also contain inflectional affixes that are restricted more-or-less to the same noun stems and verb stems as defined by their distribution under derivation.

(3) -like
    morphology: $[_{A[-0]}$ N[-0] __]
    syntax: null
    semantics: like'(N') = M(CN) "like an N"
(4) -er
    morphology: $[_{N[-0]}$ V[-0] __]
    syntax: null
    semantics: er'(V')= CN "x:[V'...(x)]"

The syntactic field in each of these lexical entries indicates that the lexeme itself is not a formative in the syntax. There is no syntactic noun -er in the sentence The farmer killed the duckling. But that is not to say that an adjective stem formed with -like, or a noun stem formed with -er, will have no syntax. Indeed, by (1), we should expect them to behave like ordinary syntactic adjectives and nouns, as is correct.

Now let us consider the semantics. As far as the formation rules of the semantics (i.e. the syntax of the semantics) is concerned, the majority of logicians would class nouns and verbs together as predicates. To be sure, there is a tendency for nouns to be one-place predicates, and verbs to have valence greater than one, but the statistics are hardly overwhelming. Gupta 1980, however, has suggested that common nouns and verbs belong to distinct combinatoric semantic classes. For Gupta both common nouns and intransitive verbs are subject to a "principle of applicability" that can be formalized as a predicate, but only nouns are also subject to a "principle of identity" such that they can be

directly involved in the formation of restricted quantifiers, as shown in LF8 in the appendix.

As before, the use of the word "noun" for those entities that function in the semantics as restrictors of abstract quantifiers is something of a pun. I will therefore call them CNs (for "common noun," following the logical tradition) to distinguish them from syntactic nouns or morphological noun stems. The other semantic categories are also given semantic symbols that differ from the corresponding syntactic or morphological categories, but here the difference is probably too great, since the spirit of the rule in (1) clearly ought to be extended to the semantics as well.

The semantic rules in the appendix to this paper state only combinatoric features of semantic representation, neglecting content entirely. But some aspects of semantic content, in particular something like the notion of thematic role, familiar to modern linguistics since Gruber 1965 and Fillmore 1968, might well require representation in an adequate autonomous-modular grammar for natural languages, as Steven Lapointe 1987, Jan Terje Faarlund (1989), and Eric Schiller (1989) have variously suggested. While each of their methods for incorporating thematic information has something to be said for it, for purposes of discussion, I will use the system argued for by Faarlund, for one thing, because it is the simplest.

Faarlund proposes that thematic roles ought to be displayed on an autonomous level, distinct from both phrase-structure syntax and combinatoric semantics, and further assumes that there are just three universal roles, AGENT, THEME, and LOCUS. As in other work on relational roles, he assumes a version of the Stratal Uniqueness Law (Perlmutter and Postal 1983), or Functional Completeness and Coherence Principles (see chapters 2 and 4 of Bresnan, ed. 1982), that rules out single assigner-role structures in which more than one item is assigned to the same role. Therefore, wherever we encounter structures with two identical roles, we must assume that there are two role-assigners in this dimension of representation (though not necessarily at any of the others.) As has been pointed out by various authors over the last twenty years, various sorts of predicative elements will differ as to the number and type of thematic roles that they assign. The verb <u>run</u> assigns an agent, <u>kill</u> an agent and a theme, <u>slip</u> a theme, and so forth.

What is striking as far as part-of-speech distinctions are concerned is this: While it is typical for verbs to assign specific relations with palpable semantic content to their arguments, it is not typical for nouns. With one principled type of exception, ordinary nouns such as <u>sparrow</u>, <u>sugar</u>, <u>table</u>, <u>liberty</u>, and <u>Stalin</u>[3], be they natural kind terms or artifacts, common or proper nouns, mass or count, concrete or abstract - all regularly fail to assign role relations of any sort. If x is a sparrow, it is not so in an agentive fashion, or a locative fashion, and is only a theme if "theme" is a catch-all category with no particular actual semantic content. The themes of the vast majority of verbs deserve the more significant name "patient" in that an aspect of their existence or location is established or altered in the event-type indicated by the verb. Even relational nouns like <u>brother</u>, <u>name</u>, or <u>hand</u>, which are two-place predicates semantically, normally do not assign ordinary thematic relations. If x and y are brothers, neither is an agent, and neither a theme: they are simply brothers. Of course <u>Stalin killed the sparrow</u> has Stalin as agent and the sparrow as theme, but that is because of the roles that the verb assigns. In the predication sparrow'(x), there is no indication whatsoever of the bird's destiny.

The exception alluded to above are those nouns that we call nominalizations. Such nouns regularly do assign standard role relations to their arguments. The obligatory arguments (i.e. the arguments that are quantified in expressions of the form <u>each N</u>, etc.) of <u>runner</u> and <u>killer</u>, are clearly agents, for example, and the obligatory argument of

discovery (in one sense) is clearly a theme. But these are not typical nouns, a point to which I will return.

Preliminarily, then, we can say that nouns are characterized by the following four potentially independent properties in four modules:

(5) a noun is:

> morphology: N[-0]
> syntax: N[0]
> semantics: CN^{-n}
> thematics: assigns no relations

Typical verbs, on the other hand, differ with respect to all of these properties.

(6) a verb is:

> morphology: V[-0]
> syntax: V[0]
> semantics: F^{-n}
> thematics: assigns relations

It is this maximal contrast between the class of nouns and the class of verbs that makes them the two classes that will be found in every part-of-speech system. The third most popular class of lexemes, that of adjectives, contrasts strongly with both verbs and nouns in three out of four components, though relationally, I can find no clear generalizations. Many adjectives, like big, blue, and wooden would seem to assign no relation, like nouns, but others, such as careful, fond, and honest do. Note also that adjectives are subsidiary to nouns in both the syntax and semantics: there can be nouns without adjectives, but not adjectives without nouns.

The typical schema for adjectives, then, is as in (7).

(7) an adjective is:

> morphology: A[-0]
> syntax: A[0]
> semantics: M(CN)
> thematics: no generalization

What is of interest here is the fact that the lists of properties in (5), (6) and (7) characterize the typical members of the group, but not every member. Lexemes may depart in one or more ways from these sets of inter-modular expectations, and be fully nominal, verbal, adjectival, etc. in one or more components, but not in all of them. We have already seen that nominalizations, though syntactically and morphologically nouns, are thematically more like verbs in that they borrow their role-assigning properties from the verbs they are derived from.

This approach to parts of speech systems raises any number of questions concerning the association of different sets of properties from different modules and only a tiny fraction of them can be taken up in a paper of this length. All that I can do in the space and time allotted to me is explore a few examples of aberrant sorts of English words in an effort to illustrate the power of the system sketched here.

Let us first consider words like <u>bitch</u>, that behave in some respects like nouns, and in others like adjectives, as discussed by McCawley (1987, 1989). Suppose that these are, as McCawley suggests, morphologically and syntactically nouns, but semantically akin to typical adjectives.

(8)   bitch (one sense):

          morphology = N[-1, -PL]
          syntax = N[0]
          semantics = M(CN) "difficult"

Since <u>bitch</u> is not a syntactic adjective and therefore will not fit in the phrase defined by SF11, *<u>a bitch problem</u> is ungrammatical. Instead what we find is <u>a bitch of a problem</u> with semantically empty elements filling in to rescue the syntax. The preposition <u>of</u> is clearly a lexeme that sometimes has no semantic relevance whatsoever, explaining its appearance in nominalizations like <u>the singing of Caruso</u>, and <u>the destruction of the city</u>. On the assumption that <u>bitch</u> is not a semantic CN, there is only semantic room for one of the two articles that we find in the syntax of <u>a bitch of a problem</u>. Since the first can be replaced by other articles, such as the emphatic indefinite in <u>one (real) bitch of a problem</u>, or demonstratives as in <u>that bitch of a problem I told you about</u>, whereas the second cannot (*<u>a bitch of one problem</u> *<u>a bitch of that problem I told you about</u>), it is clear that it is the external article that is a genuine quantifier. Many of the excluded quantifier expressions are impossible because this particular word <u>bitch</u> only occurs in singular form, a fact that is stipulated in the morphological field in (8).

Thus we arrive at the bimodular structure that McCawley suggests:

(9)
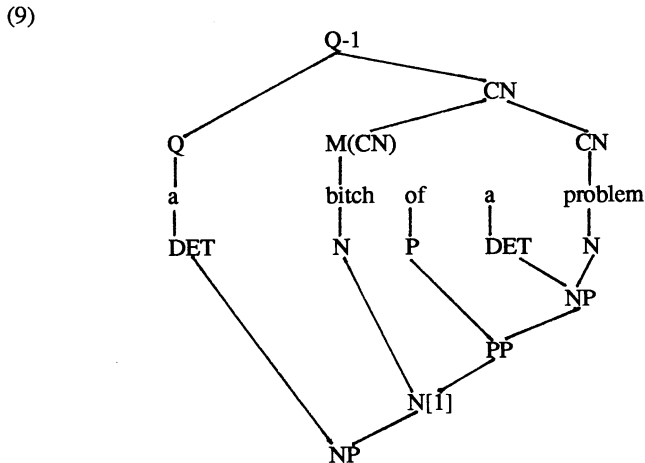
In predicate position things are a little different. Here the article that appears with <u>bitch</u> is semantically empty, and is therefore what we might call an expletive article (on analogy with semantically empty elements like <u>it</u> and <u>there</u>). It is still required by the syntax, since what follows the copula syntactically (in English, but not in German) must be a maximal projection, as shown in SF12. As a syntactic count noun (though not a semantic noun at all), the word <u>bitch</u> requires an article to form the NP required by the

syntax. Thus in the sentence <u>This problem is a bitch</u> the verb phrase has the bimodular structure given in (10), where the syntactic formative <u>a</u> is unconnected with anything in the semantics. For this reason one cannot say *<u>A problem is that bitch</u> (which is well formed on an unpleasant, sexist reading, but not with the intended meaning).

(10)
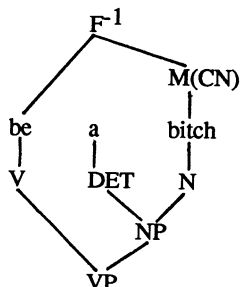
Semantic nominal modifiers that are syntactic mass nouns don't need an empty article when used predicatively, since such nouns form NPs directly. But when used as prenominal adjectives that form N[1]s, an article is still needed for both the syntax and the semantics of the whole NP. Therefore the contrast between the a. and b. examples below:

(11)   a. This problem is hell.
        b. *This problem is a hell.
(12)   a. This is a hell of a problem.
        b. *This is hell of a problem.

Let us turn to another variety of anomalous nouns, those that pretheoretically seem to function as a quantifiers, such as <u>bunch</u> and <u>lot</u> in (13).

(13) A bunch/lot of students demonstrated.

Suppose we simply say that <u>bunch</u>, and other quantificational nouns like it, have lexical entries such as (14), where the syntacto-semantic correlation between the syntax and semantics is overridden by specific lexical stipulation.

(14) bunch:

        syntax = N[0]
        semantics = Q^{-1} "many"

As a syntactic noun, <u>bunch</u> will not occur in configurations typical of quantificational adjectives or determiners (*<u>bunch boys</u>, cf. <u>many boys</u>), but rather where nouns occur, namely as heads of NPs. Once again, as with adjectival nouns like <u>bitch</u>, <u>bunch</u> will occur with the semantically empty syntactic indefinite article, and with the semantically empty preposition <u>of</u>, in exactly the same syntactic configuration as the adjectival noun: <u>a bunch of students</u>. Indeed, it occurs in the same syntactic frame as its non-quantificational homonym: <u>a bunch of flowers</u>.

It is important to note why structures like these are the only configurations possible. Both *a problem of a bitch and *boys of a bunch, while syntactically well formed, will not match the semantic structures they must have in virtue of the stipulated properties of bitch and bunch. As a modifier of CN expressions and as a quantifier, respectively, both will have semantic scope over the noun meanings they take as arguments, and therefore, all other things being equal, will have to occur higher in the syntactic tree than the nouns they occur with by the General Homomorphism Constraint on intermodular structural matching given in (1a).

Several properties of the word bunch are directly accounted for by this simple idea of its cross-categorial status, including the following:

1) Quantifier scope ambiguities: Quantificational bunch and its ilk generate scope ambiguities that do not involve existentials. In I haven't seen a bunch of flowers, the ambiguity is one of existence: either I didn't see a specific bunch of flowers, though I might have seen some, or I saw no bunch of flowers at all. But in I haven't seen a bunch of articles on that topic, either there is a specific set of articles I haven't seen, or there are only a few I have seen. In other words, there is no sense here in which an existential quantifier is inside the scope of a negative operator. On the present theory this follows from the simple fact that there is no existential quantifier in a bunch of articles. (See structure (15) below.)

2) Peculiarities of modifiers: Any meaningful element of syntax must have both coherent syntactic and semantic roles to play, that is, it must occur in syntax in positions where the syntactic categories to which it belongs can occur, and it must must behave in semantic structures in the way that its semantic categorization determines.
Now if we consider the bimodular structures like (15) in which quantificational bunch can be found, it is immediately apparent that normal adjectival modifiers of the syntactic N-bar bunch of students should be impossible.

(15)
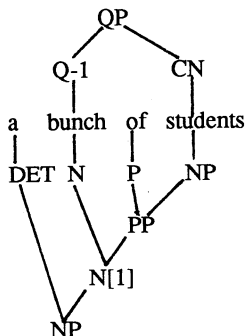
Adjectives take CNs to CNs in the semantics, but the only semantic CN in (15) is students, and if that CN is to be modified, the modifier should show up with it, by the same constraint demanding hierarchical congruence mentioned above. Interestingly, though, misplaced modifiers are sporadically acceptable in just these structures, while misplaced modifiers are totally impossible in corresponding syntactic structures with normal semantics. .

(16) an impressive bunch of publications
    = "a bunch of impressive publications"
(17) I've never met a bigger bunch of idiots.
    = "I've never met a bunch of bigger idiots."
(18) the disgusting remains of animals
    not = "the remains of disgusting animals"

While the readings in (16) and (17) are violations of (1a), we may observe that they are, in an intuitive sense, not as bad as they would be in the case of (18). Since the meaning of the modifiers in (16) and (17) could not possibly apply to the elements they are syntactically associated with, no confusion could result. Still it is puzzling why the misplacement should be allowed at all, let alone sound better than their translations into "logical order", as is the case with (17).
There is one other sort of modification allowed, as exemplified in (19):

(19) A big bunch of students marched on the Administration Building.

Here there is no violation of the general Homomorphism Constraint, since <u>big</u> in such examples is clearly a semantic modifier of the quantifier, much as <u>very</u> is in <u>very many students</u>: a big bunch of flowers can consist of four enormous flowers, a big bunch of students must contain many more students than that, regardless of their size. This <u>big</u> must be the same word observed by McCawley in examples like (20), a syntactic adjective that is semantically an intensifier of adjectival or quantifier meanings. In fact, <u>big</u> can intensify ordinary predicate meanings, as in (21) or (22). One doesn't find this use in the case of verbal examples like the ungrammatical (23) for the simple reason that <u>big</u> is always an adjective, and hence whatever it modifies must somehow or another be expressed as a noun.

(20) This problem is an even bigger bitch than that one.
(21) She is a big Cubs fan.
(22) Sam is the biggest eater in the state.
(23) *He roots big for the Cubs.

(24) big
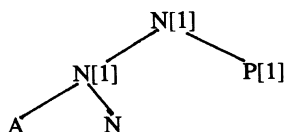
    syntax = A
    semantics = M(X) "X to a high degree"

A similar unusual adjective is <u>whole</u>, an intensifier of quantifier meanings in <u>whole lot</u> and <u>whole bunch</u>. It cannot modify the meanings of other sorts of semantic functions, however: *<u>whole bitch</u>, *<u>whole Cubs fan</u>. This highly restricted word therefore has the following lexical entry:
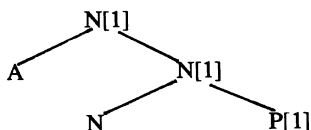
(25) whole

    syntax = A
    semantics = $M(Q^{-1})$ "very much X"

3) Limited conjoinablilty. According to the simple syntactic rules SF9 and SF11, an adjective may be combined with an N-bar either inside or outside a prepositional phrase, as in (26) and (27), respectively:

(26)

```
              N[1]
             /    \
         N[1]      P[1]
        /   \
       A     N
```

(27)

```
              N[1]
             /    \
            A      N[1]
                  /    \
                 N      P[1]
```

Which structure one finds is not an arbitrary matter, but, through (1a) depends on semantics. Thus subcategorized PPs will be inside restrictive adjectives, a result that follows in the present framework from the general homomorphism constraint plus the fact that subcategorized PPs, or rather their nominal objects, will be semantic arguments of lexical nouns (cf. LF11). Therefore the phrase <u>unemployed brother of Harry</u> will have the dual structure in (28), where the constituency of the syntactic and semantic trees match, rather than that in (29), where they do not.

(28)

```
                      N[1]
                     /    \
                   A       N[1]
                          /    \
                         N      P[1]
              unemployed brother   of Harry

                M(CN)      CN^-1          x
                                    CN
                         CN
```

(29)

```
                      N[1]
                     /    \
                 N[1]      P[1]
                /   \
               A     N
           unemployed  brother      of Harry

             M(CN)     CN^-1           x
                              CN
                      CN
```
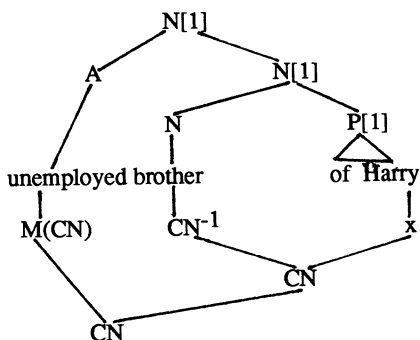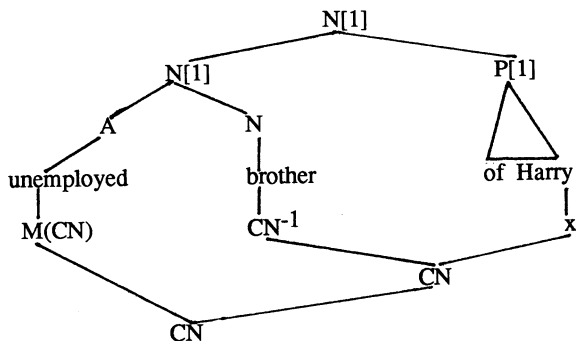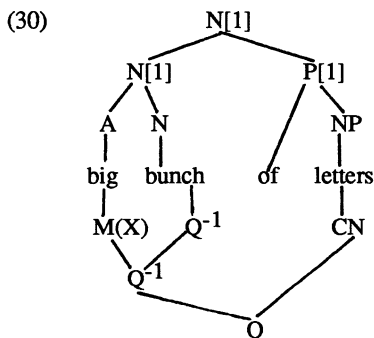
But in (30), the adjective will modify the quantificational noun in the syntax, because it takes the quantifier as an argument in the semantics.

(30)

Therefore, there will be no N-bar <u>bunch of letters</u> in (30), and the following will be ungrammatical:

(31) *I noticed a whole bunch of students and lot of professors.
    (Compare: I put in a whole bunch of celery and clove of garlic.)

4) Limited <u>one</u> pronominalization. This identity of sense pronominal needs an N-bar antecedent in syntax with a CN meaning to refer to. Thus when we have structure like <u>a picture of a dog that wrote a novel</u>, one can usually refer to any of the meanings of the included N-bars.

(32) Bill has a picture of a dog that wrote a novel,but only Fred has
    a) actually written one himself.
    (one="novel")
    b) actually met one personally.
    (one="dog that has written a novel")
    c) actually taken one.
    (one="picture of a dog that has written a novel")

But while there are two N-bars in <u>a bunch of students</u>, there is only one CN meaning, as shown in (15). Therefore:

(33) Bill knows a bunch of Republicans, but Fred is actually friends with one
    (one = "Republican"; not "bunch of Republicans")

5) Agreement. One of the most striking features of quantificational nouns is that they form noun phrases whose agreement properties are those of the object of the subordinate preposition:

(34) A lot of people *doesn't/don't think so.
(35) A lot of sugar isn't/*aren't healthy.

There is a morpho-syntactically plural alternative to   <u>a lot</u>, namely <u>lots</u>. When this alternative is chosen, the agreement pattern is unaffected:

(36) Lots of people *doesn't/don't think so.
(37) Lots of sugar isn't/*aren't healthy.

What this shows is that the agreement in these cases is controlled by semantics, rather than morpho-syntax, a phenomenon that has been documented elsewhere (Morgan 1972). However one chooses to handle semantic agreement, it is obvious that having semantic structures that make structures with quantificational head nouns identical to structures with quantifiers will make the extension to this class of phrases a fairly simple matter.

Appendix: Grammar Fragments
    1. Syntax

(SF1)   S --> N[2] V[1]
(SF2)   N[2] --> DET N[1]
(SF3)   V[1] --> V[0, SF3]
(SF4)   V[1] --> V[0, SF4] N[2]
(SF5)   V[1] --> V[0, SF5] V[1,[to]]
(SF6)   V[1] --> V[0, SF6] V[1, BSE]
(SF7)   V[1] --> V[0, SF7] S[1, IND]
(SF8)   S[1, IND] --> COMP[SF8] S[FIN]
(SF9)   N[1] ---> N[1] (PP)
(SF10) P[1] ---> P N[2]
(SF11) N[1] ---> A N[1]
(SF12) V[1] ---> V X[2]

2. Semantics

(LF1) $F = F^{-1}(i)$
(LF2) $F^{-1} = F^{-2}(i)$
(LF3) $F^{-2} = F^{-3}(i)$
(LF4) $F = O^{-1}(F)$
(LF5) $F^{-1} = O^{-2}(F)$
(LF6) $F^{-2} = O^{-3}(F)$
(LF7) $X = MX(X)$
(LF8) $Q = Q^{-1}(CN)$
(LF9) $F = Q(F)$
(LF10) $CN = M(CN)$
(LF11) $CN = CN^{-1}(i)$

3. Morphology

(MF1) $X^{-1} --> X^{-0}, Y$
(MF2) $X^{0} --> Y^{0}, X$
(MF3) $W^{-2} --> X^{-n}, Y$
(MF4) $X^{-0} --> X^{-0}, Y^{-0}$

Footnotes
[1]Tefloner is possible, of course, as the agentive nominal of a neologistic verb teflon (perhaps "to apply teflon to"), but that reinforces the point being made here, which is simply that the agentive suffix only applies to morphological verbs.
[2]For some important caveats on this analysis, see Jacobsen, 1979.
[3]Proper nouns are not ordinarily treated as predicates, but whatever they are, they are ordinarily not role assigners.

References

Ackrill, J.L. ed., 1987, <u>A New Aristotle Reader</u>, Princeton University Press, Princeton.
Bloch, Bernard, and George L. Trager, 1942, <u>Outline of Linguistic Analysis</u>, Linguistic Society of America, Baltimore.
Bresnan, Joan, ed., 1982, <u>The Mental Representation of Grammatical Relations</u>, MIT Press, Cambridge.
Faarlund, Jan Terje, 1989, "Autostructural Analysis," paper presented at the Workshop on Autolexical Syntax, University of Chicago, 1989.
Fillmore, Charles J., 1968, "The Case for Case," in Emmon Bach and Robert Harms, eds., <u>Universals in Linguistic Theory</u>, Holt, Rinehart, and Winston, New York.
Gruber, Jeffrey S., 1965, <u>Studies in Lexical Relations</u>, doctoral dissertation, MIT, Cambridge.
____, 1976, <u>Lexical Structure in Syntax and Semantics</u>, North Holland, Amsterdam.
Gupta, Anil, 1980, <u>The Logic of Common Nouns</u>, Yale University Press, New Haven.
Jacobsen, William H., 1979, "Noun and Verb in Nootkan," in Efrat, ed., <u>The Victoria Conference on Northwestern Languages</u>, 83-153, British Columbia Provincial Museum, Victoria.
Lamb, Sydney M., 1966, <u>Outline of Stratificational Grammar, Revised Edition</u>, Georgetown University Press, Washington.
Langacker, Ronald W., 1987, "Nouns and Verbs," <u>Language 63</u>, 53-94.
Lapointe, Steven G., 1987, "Some Extensions of the Autolexical Approach to Structural Mismatches,"in Geoffrey J. Huck and Almerindo E. Ojeda, eds., <u>Syntax and Semantics, Volume 20: Discontinuous Constituency</u>, 152-184, Academic Press, Orlando, Florida.
Lockwood, David G., 1972, <u>Introduction to Stratificational Linguistics</u>, Harcourt Brace Jovanovich, New York.
McCawley, James D., 1987, "A Case of Syntactic Mimicry," in Rene Dirven and Vilem Frid, eds., <u>Functionalism in Syntax</u>, 459-70, John Benjamins, Amsterdam.
McCawley, James D., 1989, Individuation in and of Syntactic Structures," chapter 6 of Baltin, Mark R., and Anthony S. Kroch eds.,1989, <u>Alternative Conceptions of Phrase Structure</u>, University of Chicago Press, Chicago.
Morgan, J.L., 1972, "Verb Agreement as a Rule of English," CLS 8, 278-286.
Perlmutter, David M., and Paul M. Postal, 1983, "Some Proposed Laws of Basic Clause Structure," in David M. Perlmutter, ed., <u>Studies in Relational Grammar 1</u>, 81-128, University of Chicago Press, Chicago.
Sadock, Jerrold M., forthcoming, <u>Autolexical Syntax: A Theory of Parallel Grammatical Representations</u>, University of Chicago Press, Chicago.
Sapir, Edward, 1921 (1949), <u>Language: An Introduction to the Study of Speech</u>, Harcourt, Brace, and World, New York.
Schachter, Paul, 1985, "Parts of Speech Systems," Chapter 1 of T. Shopen, ed., <u>Language Typology and Syntactic Description, Volume I: Clause Structure</u>, Cambridge University Press, Cambridge.
Schiller, Eric, 1989, "The Case for Autolexical Case," paper presented at the Workshop on Autolexical Syntax, University of Chicago, April 16, 1989.
Swadesh, Morris, 1939, "Nootka Internal Syntax," <u>IJAL 9</u>, 77-102.

From Tone to Stress:
Mechanisms and Motivations[1]

Joe Salmons
Purdue University

0. Introduction

Tonogenesis has established itself as a standard topic among historical linguists working with tone languages.[2] The other side of the same coin, i.e. how tonal systems become stress-accent (or intermediately, pitch-accent) systems, has barely been treated, although numerous scholars have expressed interest in this issue, and a few have treated specific instances of tone-to-stress shift, including Goldsmith, Li, and others. Beckman (1986: 32) is perhaps most explicit about it: "There are as yet no theories of the genesis of accent so well substantiated as is the now standard theory of tonogenesis."

Examples of such tone loss and accentual genesis are well attested across southeast Asia, East and West Africa and in numerous contact languages from around the world. This paper briefly treats structural as well as sociolinguistic and language-contact aspects of tonal loss. Only a holistic explanation, namely one including both well-developed language-internal and language-external arguments, can comfortably account for this cross-linguistically very common tendency. In particular, I focus on the crucial transitional stage away from tone to stress, especially the more specific problem of pitch-to-stress shift, and note some possible external motivations for such a transition. The internal mechanisms by which the shift takes place must be understood, but the motivations must be sought language-externally. The crucial argument here is that tone-to-stress shifts around the world show profound parallels and that the differences among such developments are at least sometimes illusory or superficial.   The link between internal and external considerations is the role of intonation in word-level accentuation and accentual change.

The recognition of a combination of both internal and external considerations as absolutely necessary for a complete understanding of change has become commonplace among theoreticians of language change. This basic tenet is, for example, agreed on by scholars with such diverse theoretical perspectives as Kiparsky (e.g. 1984, 1988), Thomason and Kaufman (1988), and Anttila (1989).[3] As Kiparsky has argued (1988: 375), a theory of language change is not an autonomous theory, but must incorporate a broad range of other considerations. While this is recognized in theoretical treatises, such integrated explanations remain relatively rarer in specific case studies. In the case at hand, internal and external motivations both contribute to a particular type of language change. The changes treated represent simplification within the linguistic system and these changes also tend to take place under sociocultural circumstances favoring language change in general and accentual change in particular. This article tries to bridge the internal-external gap in practice, although limitations of data and frameworks leave matters vague or speculative on several crucial points. In that sense, this piece tries to point to some areas where further work is needed.

1. Internal aspects

The developmental path toward stress appears often to involve a series of steps, including these three:

(1) Tone
(2) Pitch accent
(3) Stress (fixed)

This sequence of changes can be precisely and concisely described using nonlinear frameworks, and recent work in phonetics lends empirical support to nonlinear descriptions. See for example Beckman (1986), who, however, proposes a single fundamental distinction between stress and non-stress accent.[4]

Of these three classical accentual types, the first and last have been relatively straightforwardly distinguished in numerous handbooks. One of the most interesting points of contrast, one developed below, is Beckman's claim that tone functions more like a distinctive feature—being, for example, susceptible to neutralization in certain contexts—while accent functions more as an "organizational feature" (1986: 2). That is, accent creates a focus within a string of speech, by giving prominence to certain areas and denying it to others.

Pitch accent is where the action is, since it covers a broad continuum of specific accentual systems, as noted by van der Hulst and Smith (1988: ix). The clarification of the nature of pitch accent is indispensable for an understanding of its role as a transitional stage in tone-to-stress shift. While the role of stress in pure tone languages is at best unclear, and while pure stress languages can be described without reference to tone, both tone and stress are necessary elements in understanding a pitch-accent system. Along these lines, three types of pitch accent are distinguished by Inkelas and Zec (1988), based on the interaction between tone and stress:[5]

(1)   only tone needs to be marked in the lexicon and stress can be predicted from tone, e.g., Serbo-Croatian;
(2)   only stress needs to be marked to predict tone, for example in Norwegian;
(3)   both tone and stress must be marked and they function independently, as in Japanese.

In this paper, I divide pitch-accent systems into those three categories: languages in which stress is basic (with tone predictable from it), those languages in which tone is basic (and stress is predictable from it), and those languages in which both seem to be required underlyingly.[6] Such a distinction is necessary to make sense of the extremely broad notion of "pitch-accent systems".

Given pitch accent as a transitional stage from tone to stress, an uncontroversial point to my knowledge, the question becomes one of transitions between types of pitch-accent systems. That is, do languages progress uniformly from one to another of these three types of pitch accent on their way toward stress? Van der Hulst and Smith argue that there is no universal path of development from pitch accent toward stress. They contrast the developments posited for Bantu with those in Otomanguean languages, summarizing a number of studies contained in their volume (1988: xxi-xxii). In Bantu, trends include restrictions on tone patterns that can occur, replacement of the H-L distinction with H-Ø (i.e. the development of a privative distinction between tone and no tone), and development of metrical structure with the designated element corresponding to H. Such stages would appear to fall within the tone-based pitch-accent system in the taxonomy above. In Otomanguean, they suggest, following Hollenbach (1988), that a word stress system leads to reduction of tonal possibilities in unstressed syllables and, sometimes, to a shift of the tone pattern onto the stressed syllable. This would appear to represent a stress-based pitch-accent type.

However, this distinction seems illusory for a number of reasons. As noted above, pitch-accent systems, like the two examples just mentioned, must be described with reference to both tone and stress. In the case of Bantu languages, van der Hulst and Smith describe the evolution with reference to tone, and in the case of Otomanguean, they trace the role of stress. This would seem to imply that tone remains basic longer in Bantu during tone loss, while in Otomanguean stress becomes basic earlier in the process. In fact, although they assign the rise of metrical structure a late role in Bantu, it is unclear how the tonal reductions would take place without interaction with metrical prominence. In order for the Bantu account to differ significantly from the Otomanguean, the relative chronology

sketched for Bantu would have to imply that a H-L contrast might be lost in a given position without any reference to metrical prominence (or stress, to keep the more specific formulation).[7] Likewise, it would imply that a H-L contrast might, at the same time, be maintained in a position that would not be prominent. However, Goldsmith (1987a) specifically ascribes tonal developments in Kirundi and Kinyarwanda to the "rise of rhythmic structure" in those languages, something to which we will return below. But more directly contradictory evidence is also available from Bantu.

At least some Bantu languages had various kinds of fixed-stress systems. A stress shift in such a language can apparently lead to tone shift, as argued for Kikuyu by Clements and Ford (1979). In this case, Kikuyu has shifted stress one syllable to the right and stress has essentially taken the H with it, rather than spreading rightward as in a system like that of Kikamba:

Proto-Bantu	Kikuyu	Kikamba/Tharaka	
*-yéné	-ēnέ	-έnέ	'to belong to
I /	I I	I /	someone else'
I/	I I	I/	
I/	I I	I/	
H H	L H	H H	

That is, stress is implied to be basic within the pitch-accent system at the point when stress-shift takes place, as the H assignment is based on stress. This Bantu example is thus not at all in conflict with the chronology of Hollenbach (1988) for Otomanguean, nor are other instances of tone-to-stress change that I know of. At most, the difference between the Otomanguean example and the Bantu examples might be a difference in how metrical prominence is realized, *viz.* stress versus pitch. We would see then a general tendency for metrical prominence to creep into tonal systems, eventually eroding them in favor of a stress-accent system.

A very brief look at Bantu lends support to this as a general underlying principle in Bantu accentual development. Carter (1973) gives an overview of Bantu tone which can, for present purposes, be reduced to a typology consisting of four categories: (1) loss of tonal distinction outside of stem-initial syllables, (2) loss of tonal distinction within stem-initial syllables, (3) "various limits on tonal distinctions in stems", and (4) tone loss. This taxonomy can be understood as reflecting several diachronic layers in the development away from the Proto-Bantu two-tone system. The first two types show, respectively, a H-L distinction in the first syllable of stems (with loss of distinction in the second) and loss of distinction in first syllables in stems, e.g. Herero, which maintains first syllable tone distinction and Nyankore, which loses it in first syllables.[8] In short, these would reflect a single more general process and the sole significant difference rests on whether the first or second syllable of the verbal stem becomes metrically prominent, in line with Goldsmith's (1987a and elsewhere) arguments that an alternating metrical prominence has brought about shift of H tones to associate with such metrically prominent positions. The fourth type is complete tonal loss, found particularly in the eastern Bantu area, and would represent a stage of further development, i.e. tonal reduction, from the two just-mentioned types. That leaves only the very vague category of "various limits on tonal distinctions in stems", at least some of which I think can be dealt with in similar terms, although space does not permit discussion here.

If this general path of tone-to-stress evolution is accepted, at least for the moment, the real question then becomes exactly where the transition from a pitch-accent system to a stress-accent system takes place within this evolution.[9] A number of possibilities present themselves, for instance:

(1) A limit of one H per word;
(2) Development of a prominence system (an "organizational" function);
(3) Loss of distinctions outside of the one crucial syllable.

Shih (1985) seems to suggest that steps 2 and 3 follow from what Shih calls a High Pitch Constraint (like 1) in Fuzhou, limiting words to a single high tone per word. That is, once a word can have only one high tone, that syllable automatically becomes prominent. That prominence is already an accentual rather than tonal characteristic. This is in line with Beckman's understanding of accent, noted above, where it functions as an organizing principle, marking place in various domains, from the word (or even smaller units) to the phrase and beyond. The development of metrical prominence would also then lead to the reduction of contrasts elsewhere, since prominence is by definition relative.

Along similar lines, Goldsmith (1987a: 76-77) suggests the development of prosody based on H tone—"an inherently prominent syllable"—becoming an "anchor" for metrical structure. He goes on to find two factors likely to play a "causal" role in the rise of metrical structure, either a limit of one H per word, presumably by creating a possible anchor for metrical structure, or the rhythm established in languages like Kirundi by a tonal pattern of HLHL in verbs with two objects attached. The system Goldsmith treats would clearly be a pitch-accent system with tone underlying, i.e. with H anchoring the prominence. Returning now for a moment to the question of precisely defining where the transition from pitch to stress comes, one might hypothesize that the loss of Goldsmith's "anchor" would define that shift to stress. When the anchoring of metrical prominence in terms of a tone is lost, the system would be describable purely in terms of stress.

In a different formulation of matters, Goldsmith (1988: 87) sees the location of metrical prominence within the word as determining the syllable to which the H is attracted. That is, final prominence would pull a H to final position, as opposed to what Goldsmith calls "middle prominence" in Kirundi, which results in H tones in the middle of words. This would imply the reverse order—rather than the position of a single H allowing for or inducing the rise of metrical structure, here metrical structure would determine where the H occurs. These formulations seem to be at odds with one another, since the second would have to involve prior development of metrical prominence conflicting with the placement of the inherently prominent syllable, namely the syllable bearing a high tone. This must presume coexisting but independent stress and tone patterns at an earlier stage of Bantu. Otherwise, we again have the problem noted above of how a metrical structure could *arise* in conflict with H tones, i.e. inherently prominent syllables. If we go back to assuming a preexisting pattern of metrical prominence, this would leave us back with Clements and Ford's Kikuyu situation, where a shift of stress mandates a shift of tone. Within the context of Bantu, Goldsmith's second treatment of matters—with Hs being attracted into metrically prominent positions—would seem a more likely secondary development, rather than an original motivation for tonal structure.

Another way of looking at this transition can be developed. Yip (1988) argues that the Obligatory Contour Principle—forbidding underlying sequences of the same melodic feature and mandating spread in the case of tone—is at work in stress-accent systems as well as in pitch or tone systems. But the concrete manifestations of the OCP would be different in the two systems. Namely, spreading is by its very nature absent in stress-accent systems, in which the OCP is used instead to avoid "stress clash" (Prince 1983), rather than resulting in tone spread.

When the OCP functions to avoid stress clash rather than to bring about tone spreading, the transition from pitch to stress would seem necessarily complete. Presumably, for example, Serbo-Croatian would be considered a kind of stress-accent language if there were no tone spreading.[10] This would seem to be the latest stage at which tone could play a significant role; that is, it would still be possible for tone to be lost, taking spreading with it.

The rise of a stress-like pattern of prominence might prove a crucial step before that one.[11] Since the OCP (which was originally used for tone languages) has been extended by Yip and others to stress-accent systems, it is worth asking whether the OCP can be applied productively to the understanding of pitch-accent systems. I would argue that it can be, in the following way. The OCP avoids underlying sequences of the same tone on the

one hand and prevents "stress clash" on the other. Tone and stress are understood as subcategories of a more general category, accentual *prominence* (Prince 1983: 88-89). While stress-accent and tone languages utilize one or the other type of prominence in a fundamentally more significant way than the other, pitch-accent languages utilize both in important ways. So, we might expect some preference for avoiding prominence clash, which would be one level more abstract than sequences of identical tone or stress clash. That is, the OCP would avoid the occurrence of sequences of prominent syllables. For this reason, changes like Goldsmith's "tone-accent attraction condition" (1987b) become understandable in a broader framework.

So far, I have suggested some places where one could draw distinctions among tone, pitch and stress as stages in tone-to-stress shifts and have shown that the paths away from tone toward stress show profound similarities across a number of languages and families in Middle America, East Africa and Asia. The above suggests a generally uniform path from tone to stress accent, one which captures diachronic processes from diverse languages within a single description. The differences are superficial here, e.g. whether the first or second syllable of a stem occupies the metrically prominent position and thus attracts the H, retains tonal distinction, and ultimately in some languages develops stress.

How can this change be motivated internally? The traditional approach is to show that a change from tone to stress represents simplification, since, for example, rather than requiring the marking of each syllable for a particular tone in the lexicon, a stress language needs at most one syllable per word marked for primary stress. See Clements and Goldsmith (1984) and Salmons (in press) for more detailed arguments on simplification. In the transitional stage, a pitch-accent system, the issue of simplification can become clouded. A rule restricting words to no more than one H might, under certain circumstances, move the language from being a tone language to being a pitch-accent language and it would simplify the underlying representation. However, if both tone and stress must be understood independently, the rise of stress would represent a clear complication of a previously tonal system. Generally, though, the changes can be understood as simplification, so I will not pursue the matter here.

2. External aspects
Internal descriptions, however formally elegant, leave us with incomplete motivations for the change from tone to stress. Clements and Goldsmith ask why some languages remain tonal over millennia while other related languages become stress-accent languages. This seems to imply that such explanations could be sought language-internally without consideration to external factors. This implication means that their question ultimately proves ill-formulated from a historical linguist's point of view. That is, the underlying causes of language change are not to be found purely within the language, but rather in language use by a community. Kiparsky (1988: 375) gives us a formulation of the matter even more directly relevant to Clements and Goldsmith's question (noted above): "what needs to be explained by social factors is not why language changes but why change is sometimes impeded." Our understanding of the role of social factors in language change is obviously not well enough developed to do that with precision for an area as complex and ill-understood as tone and accent. Still, it appears that we have little choice but to try. Goldsmith (1987a: 77-78) uses "cause" in a restricted way, dealing with motivations that can perhaps explain why change takes place but not why it is impeded. But I would like to suggest in the rest of this paper that Goldsmith's solution to the problem of Bantu accentuation, namely the role of prosody and/or metrical prominence, fits very neatly with external motivations for language change in attested settings. That is, prosody is particularly susceptible to sociolinguistic pressure.

A powerful correlation exists between tone-to-stress restructuring and both language-contact change and *Sprachbund*-phenomena, as argued at considerable length in Salmons (in press), where detailed bibliography is also provided for the cases noted only in passing here. Perhaps the most notable attempt to place accentual change into a general theory of language-contact change is that of Thomason and Kaufman (1988: 74-75). They assign the

borrowing of "stress rules" to category 3 of their scale of language-contact related change, as part of the category of "slight structural borrowing". Their examples of such settings are immigrant languages in the United States and some Spanish-American Indian contact settings in Mexico. While most instances of accentual change appear to fall into this category, accentual change can also occur with little other structural borrowing, i.e. at or near the beginning of structural borrowing.

Language contact with tone-to-stress shift is attested in many pidgins and creoles and many African contact languages, areas of research where this has virtually become a cliché. But the evidence extends to other parts of the world with far less intense contact than found in pidginization and creolization settings. For example, Scandinavian pitch-accent systems have regularly become stress systems in relatively moderate contact settings, e.g. in Swedish dialects spoken in Finland as well as Scandinavian languages in the United States. The development of areal accentual patterns has been shown for several regions of Eurasia and Africa. Tonal areas are, of course, well established for both West Africa and Southeast Asia, with the tonal versus non-tonal isoglosses cutting in both cases directly across several language families. That is, tonal areas exist in those areas including languages of diverse genetic affiliation where related languages outside the tonal area are non-tonal. There are also indications of areal distribution of particular (sub-)types of tone systems within those more general tonal areas (cf. especially Weidert 1981: 216-219).

Stress shows similar powerful tendencies to occur areally. To give only one example, Hyman (1977) has called attention to an extensive final stress area in westernmost Asia, including Armenian, several Iranian languages, and a number of Turkic languages. An areal diffusion of fixed initial-stress accent has also been argued for prehistoric northern Europe (cf. Salmons in press), involving several subgroups of Indo-European (Celtic, Germanic, probably Italic) as well as Finno-Ugric languages.[12] This tendency also includes areas characterized by pitch-accent systems, such as part of northern California, with Hokan, Athapaskan and Algonquian languages, although this case requires further exploration. Given this clear propensity toward areal diffusion and development in conjunction with language contact, accentual restructuring would be expected to correlate with some external motivation for change. I will return to pursue this in a moment.

In spite of such powerful crosslinguistic tendencies as those just noted, Clements and Goldsmith (1984) find tone loss across the Bantu languages "unlikely [to be] a regional or diffusional phenomenon". This claim is particularly surprising in light of the maps provided by Carter (1973) and Guthrie (1967-1971), showing clear geographical patterns of accentuation across the entire Bantu-speaking area. It rests, in part, on the assumption that the outcome of areal diffusion would be identical across all the languages to which a feature or system has spread. It would have to be shown that language-contact change implied development of identical systems. This is not necessarily the case and in fact runs against the grain of much recent work on language contact. The most probable immediate motivation of tone loss—the rise of a system of metrical prominence—could be the same across at least much of Bantu but could easily be realized in superficially different ways, as seen above. That is, many of the differences across Bantu accentual systems are probably attributable to a few variables, which, as Carter (1973) notes, were simply not obvious at the time. Again though, the broad areal patterns of retention, reduction, and loss of tone are clear from such maps.

Heine (1973) correlates lack of tone in many Bantu languages with language contact, either pidginization or partial pidginization. In light of the extremely complex language contact settings in Africa, however, individual cases are not always so easily explained. Polomé (p.c.) finds substratal influence the most likely cause of tone loss in Tumbuka and some other Bantu languages, but our knowledge of the prehistorical linguistic situation is severely limited. Presumably both social factors, such as ratio of substrate versus superstrate speakers, and purely linguistic factors, such as accentual differences in substrate languages, were important in the loss of tone in Tumbuka and in the preservation of tone in Zulu and Xhosa. The social factors that encourage or discourage language change in general are, if anything, more important in accentual change, for reasons

discussed below. Unfortunately, when the sociocultural and historical background cannot be established, the ways and means of accentual change are probably lost as well. There is, I think, nothing ad hoc about this explanation—in cases where we have the data, the explanation works well, but where we do not have the data, we do not have a firm explanation and are forced to reach for more speculative solutions or to forego answering such questions.

An external (or more specifically, a language-contact or areal) motivation would fit well with the often-expressed view that prosody can lead to the appearance of new accentual patterns, i.e. that sentence-level accent can lead to alteration of word-level accent. Hyman (1977, 1978), for example, conjectures that intonation can lead to new accentual systems. Similar views for other languages can be found as well, cf. e.g. Liberman on Scandinavian (1982: 124-128) as well as the discussion by Beckman (1986: 31-36). Prosody seems in fact to be relatively susceptible to Labovian-type prestige change (or for that matter, retention based on sociolinguistic considerations). Mackey (1962) long ago declared intonation particularly persistent and particularly subtle in interference. This has been argued at least as far back as Kent (1932: 66) who attributes a reintroduction of pitch accent into Latin to Greek influence on the most highly educated classes. But intonational interference can arise in other ways, for example as a substratal feature.

Note how well this matches up with Goldsmith's several discussions of prosodic phenomena in Bantu and Hyman's attribution of stress development based on intonation. Goldsmith sees prosody as crucial to the evolution of non-tonal accentuation in Bantu. The susceptibility of prosody to sociolinguistic pressure would provide a language-external link in such changes. The "causal" factors discussed in Section 1 above still do not motivate tone-to-stress changes in some languages and cannot completely explain the retention of tone in others. Tracing the shift back to intonational considerations leaves unclear why intonation would have changed. More detailed understanding of how intonational change interacts with (word-level) accentual change is unfortunately virtually impossible given the current lack of data about intonational change. Hock's recent and extensive guide to the study of language change (1986), for example, does not deal with prosodic or intonational change in any significant way, reflecting a serious gap in the state of the art in historical linguistics today. Pierrehumbert and Beckman (1988: 239 and elsewhere) note that even the description of intonation in tone languages—something obviously very necessary for progress on the issues raised here—is still in its infancy.

3. Summary and conclusion

Tone-to-stress shift is a well attested pattern of linguistic change cross-linguistically, one which shows important parallels across many languages of the world. Even details of tone-to-stress change appear cross-linguistically consistent. While tone loss may represent simplification, internal motivations are clearly insufficient to trigger or to inhibit such a change. Such triggering and inhibiting factors must be sought elsewhere, in this case in language contact or areal diffusion. Thus, accentual change, a crosslinguistically very common type of language change, is one made especially complex because it rests not only on internal features, but also correlates clearly with external (socio-historical) considerations.

As noted at the outset, the attempt to integrate internal and external considerations into understanding language change is necessary and useful in case studies. For accentual change from tone to stress (or to some type of pitch accent as an intermediate stage), the internal and external factors both point in the same direction, simplification at work internally and features prone to sociolinguistically motivated change at work externally. This in no way constitutes any kind of attack on those who have concentrated on internal considerations (Goldsmith, Clements, Beckman, among others), but on the contrary merely attempts to put their work into a broader context. It is of course natural for those primarily interested in phonological theory not to dwell on the various external considerations necessary to understand language change; that presumably remains the work of the historical linguist. Purely internal motivations remain however ultimately incomplete

without connection to external considerations. This is true especially by omitting not simply motivations for change, but by omitting motivations for retention, following Kiparsky as cited in the introduction. Further progress in understanding the details of the role of intonation in accentual change will obviously be dependent on progress in understanding intonational change.

NOTES

[1]I would like to thank the following people for various and sundry comments on this paper and discussions on this topic: Alicja Gorecka, Monica Macaulay, Steve Wilson, and Ronnie Wilbur, as well as Greg Humpa and Dan Nutzel. All the usual disclaimers apply.

[2]These include for example the works by Hyman, Matisoff, and Weidert listed in the bibliography.

[3]There are notable exceptions here, such as Lightfoot (e.g., 1988), who essentially dismisses external considerations out of hand as at best uninteresting.

[4]I avoid the conflicts among the competing nonlinear theories and formalism as far as possible here, since such matters are tangential at best to the arguments I pursue here.

[5]A number of taxonomies and typologies of pitch accent have been proposed of late besides the one of Inkelas and Zec. Cf. a number of the contributions to van der Hulst and Smith (1988).

[6]I leave aside here traditional ways of distinguishing tone languages such as minimal pairs involving tone.

[7]The only situation I can imagine in which a single H per word could conflict with a developing prominence system would be a superstrate-substrate contact setting, where a rhythmic structure would be imposed onto a native tonal pattern. Language contact of course can and does lead to such systemic conflicts in some settings, but no one has, as far as I know, ever even speculated about such for the Bantu cases at hand here.

[8]Presumably the differences between types 1 and 2 could be reflected in stress placement, e.g. with type 1 showing stress on the first syllable of the stem, assuming a left-headed system for the moment.

[9]I consider the systems discussed in this section basically pitch-accent systems, i.e. Tonga (Goldsmith), Fuzhou (Shih), Kikuyu (Clements and Ford), rather than pure tone languages. The distinction between tone and pitch accent is a very different and perhaps more difficult topic.

[10]The result might be a non-fixed stress-accent system in the case of Serbo-Croatian. An interesting question not treated in this paper would differences in the rise of fixed versus non-fixed stress systems from pitch or tone systems.

[11]In Japanese for instance, Prince (1983: 89) says that no pattern of metrical prominence (strong versus weak) exists, putting it a step farther from a stress-accent system than Serbo-Croatian.

[12]It might also be noted that other accent-like phonological features show areal distribution, such as stød in the Baltic, preaspiration in northernmost Europe and ejectives (often occurring in stressed positions) in some parts of North America.

REFERENCES

Anttila, Raimo. 1989. *Historical and Comparative Linguistics*. Amsterdam: Benjamins. 2nd revised edition.

Beckman, Mary E. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.

Carter, Hazel. 1973. Tonal Data in *Comparative Bantu*. *African Language Studies*. 14. 36-52.

Clements, N. G., and Kevin Ford. 1979. Kikuyu Tone Shift and its Synchronic Consequences. *Linguistic Inquiry*. 10(2). 179-210.

Clements, N.G., and John Goldsmith. 1984. Introduction. *Autosegmental Studies in Bantu Tone*, eds. N. Clements and J. Goldsmith. Dordrecht: Foris. 1-17.

Goldsmith, John. 1984. Tone and Accent in Tonga. In *Autosegmental Studies in Bantu Tone*, eds. George N. Clements and John Goldsmith. Dordrecht: Foris. 19-51.

_____. 1987a. The rise of rhythmic structure in Bantu. *Phonologica 1984*, eds. Wolfgang Dressler et al. London: Cambridge University Press. 65-78.

_____. 1987b. Tone and accent, and getting the two together. *BLS*, 13. 88-104.

_____. 1988. Prosodic trends in Bantu languages. *Autosegmental Studies on Pitch Accent*, eds. Harry van der Hulst and Norval Smith. Dordrecht: Foris. 81-94.

Guthrie, M. 1967-1971. *Comparative Bantu*, 4 volumes. Farnborough: Gregg International.

Heine, Bernd. 1973. *Pidgin-Sprachen im Bantu-Bereich*. Kölner Beiträge zur Afrikanistik. Berlin: Reimer.

Hock, Hans Heinrich. 1986. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.

Hollenbach, Barbara E. 1988. The asymmetrical distribution of tone in Copala Trique. *Autosegmental Studies on Pitch Accent*, eds. Harry van der Hulst and Norval Smith. Dordrecht: Foris. 167-182.

Hyman, Larry. 1977. On the Nature of Linguistic Stress. *Studies in Stress and Accent*. Southern California Occasional Papers in Linguistics, 4. (Papers from Stréssfèst), ed. Larry Hyman. 37-82.

_____. 1978. Historical Tonology. *Tone: A linguistic study*, ed. Victoria Fromkin. New York: Academic Press. 257-270.

Inkelas, Sharon and Draga Zec. 1988. Serbo-Croatian Pitch Accent: The interaction of tone, stress, and intonation. *Language*. 64. 227-248.

Kent, Roland G. 1932. *The Sounds of Latin: A descriptive and historical phonology*. Language Monographs, 12. Baltimore, Md: Waverly Press.

Kiparsky, Paul. 1980. Concluding Statement. *Papers from the Fourth International Conference on Historical Linguistics*, eds. Elizabeth C. Traugott, Rebecca LaBrum, and Susan Shepherd. Amsterdam: Benjamins. 409-418.

_____. 1988. Phonological Change. *Linguistics: The Cambridge Survey, vol 1.*, ed. Frederick Newmeyer. Cambridge: Cambridge University Press. 363-415.

Li, Charles N. 1986. The rise and fall of tones through diffusion. *BLS*, 12. 173-85.

Liberman, Anatoly. 1982. *Germanic Accentology: The Scandinavian Languages*. Minneapolis: University of Minnesota Press.

Lightfoot, David. 1988. Syntactic change. *Linguistics: The Cambridge Survey, vol 1*, ed. Frederick Newmeyer. Cambridge: Cambridge University Press. 303-322.

Mackey, W. F. 1962. The description of bilingualism. *Canadian Journal of Linguistics* 7. 51-85.

Matisoff, James A. 1973. Tonogenesis in South-east Asia. *Consonant Types and Tone*, ed. Larry Hyman. Los Angeles: Linguistics Program, University of Southern California. 71-96.

Pierrehumbert, Janet B., and Mary E. Beckman. 1988. *Japanese Tone Structure*. Cambridge, Mass.: MIT Press.

Prince, Alan S. 1983. Relating to the grid. *Linguistic Inquiry* 14. 19-100.

Salmons, Joe. In press. *Accentual Change and Language Contact*. Stanford, Calif.: Stanford University Press.

Shih, Chi-lin. 1985. From Tonal to Accentual: Fuzhou Tone Sandhi Revisited. *BLS*, 11. 316-326.

Thomason, Sarah G., and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley, Calif.: University of California Press.

van der Hulst, Harry, and Norval Smith, eds. 1988. Introduction. *Autosegmental Studies on Pitch Accent*. Dordrecht: Foris. xi-xxiv.

Weidert, Alfons. 1981. *Tonologie: Ergebnisse, Analysen, Vermutungen.* Tübingen: Max Niemeyer. Linguistische Arbeiten, 105.

Wright, Martha. 1988. A level-based model for pitch accent languages. *Autosegmental Studies on Pitch Accent*, eds. Harry van der Hulst and Norval Smith. Dordrecht: Foris. 295-316.

Yip, Moira. 1988. The obligatory contour principle and phonological rules: a loss of identity. *Linguistic Inquiry* 19(1). 65-100.

Defined and Neutral Categories in Lexicogrammatical Patterns: The Motivated Alignment Hypothesis*

Linda Schwartz
Indiana University

1 STRONG SEMANTIC ALIGNMENT. Any general account of natural language syntax must deal with the interface of thematic structure and syntactic representation. One hypothesis regarding thematic-syntactic linking is the Strong Semantic Alignment Hypothesis, which holds that the information present in thematic structure is sufficient to determine syntactic representations or range of potential syntactic representations. The Universal Alignment Hypothesis of Relational Grammar (Perlmutter and Postal 1984) is a theory-specific statement of Strong Semantic Alignment. This paper examines the structure of lexical split predicate systems, where the relevant predicates of a language are divided into classes but where the relationship between the classes appears to be partially semantically predictable, but not wholly so. For this study, only binary splits are considered. Two much-studied distinctions are chosen: SPLIT INTRANSITIVE and INVERSION systems.[1] The systems I am concerned with are not fluid predicate systems (in the sense of Dixon 1979), and they are relatively heavily lexicalized, more so than the Korean agentive and experiential constructions examined in Chun and Zubin (this volume) for example.

It should be noted at the outset that what I am considering here is a LAST RESORT ANALYSIS, in the sense that whether one holds that partial predictability exists or whether one rejects this, there is no way to prove partial predictability, other than the absence of a more revealing account, and the only strategy by which one might hypothesize that a phenomenon is partially but not fully predictable is by trying plausible predictive principles derived from the investigation of similar systems and determining that they don't work. It can never be ruled out in principle that a more revealing and predictable account of the phenomenon will not be forthcoming in the future. However, the main point I want to make is that even if the existence of only partially predictable systems must ultimately be accepted, the investigation presented here suggests that the relation among the categories in such a system is still highly constrained.

2 CLASSICAL CATEGORY SYSTEMS. Strong Semantic Alignment defines only classical categories as discussed in Lakoff 1987, since it explicitly assumes that the principles of thematic-grammatical mapping are PREDICTIVE, in that they specify a set of necessary and sufficient conditions for the classification of any given individual predicate. Some lexical categories in split predicate systems may be defined by necessary and sufficient conditions. For example, Van Valin (to appear) provides a characterization of split intransitivity in Italian auxiliary selection where it is claimed that all verbs whose lexical semantics contain a stative predicate select the auxiliary *essere* 'to be', while all other verbs select *avere* 'to have'. Thus, the presence of a stative

predicate in the lexical semantics of a verb provides a necessary and sufficient condition for the selection of *essere* , and the complement category selects *avere* . Other analyses of split predicate systems which may be consistent with Strong Semantic Alignment in the sense of being predictive are Grimshaw's 1989 analysis of the *frighten* class of psych predicates in English, Durie's 1987 analysis of split intransitives in Acehnese, and other split predicate systems examined in Van Valin (to appear). These are the kinds of systems that Strong Semantic Alignment would lead one to expect. Now, I'll turn to an examination of some systems which seem to be crucially different from these.

3 DEFINED AND NEUTRAL CATEGORY SYSTEMS

3.1 DAKOTA SPLIT INTRANSITIVES. In the Dakota split intransitive system, there is a closed category of intransitive predicates which take pronouns parallel to those of the subjects of transitive predicates -- using the terminology of Merlan 1985, these are the subject-inflecting predicates. All predicates in this category necessarily select animate subjects. There is no evidence that volitionality, controllability, or aspect, all properties associated with subject-oriented (unergative) intransitives, are defining properties of this class. Boas and Deloria 1941 list about 90 predicates and two classes of predicates which are in the subject-inflecting category. Sample predicates are given in (1).

1) DAKOTA SPLIT INTRANSITIVES: SUBJECT-INFLECTING PREDICATES
(first person singular: *wa-* ; second person singular: *ya-*)
wačí 'dance', *psíča* 'jump', *mani'* 'walk', *slohã* 'crawl', *ʔũ* 'be (animate)', *pšá* 'sneeze', *ní* 'live, be alive', *ksápa* 'be wise', *thí* 'to dwell', *čéya* 'to weep', *léža* 'to urinate', *šičá-howàya* 'to groan with misery', *kažó* 'have diarrhea', *ɣópa* 'snore', *niyá* 'to breathe', *núni* 'to lose one's way', *blo-káska* 'to hiccough', *káɣa* 'to choke on something', all verbs of coming and going, all verbs ending in *-pha*

The other, much larger, intransitive predicate category takes pronouns parallel to those functioning as objects of transitive predicates; this category places no selectional restriction on its subject and contains predicates which select animate subjects, predicates which select inanimate subjects, and predicates which allow either. Sample predicates from this class are given in (2).

2) DAKOTA SPLIT INTRANSITIVES: OBJECT-INFLECTING PREDICATES
(first person singular: *ma-* ; second person singular: *ni-*)
t'e 'die', *c'ãzé* 'be angry', *kʰúža* 'to be ill', *bléza* 'to be sane', *čʰepá* 'to be fat', *ilé* 'burn', *sut'ũ* 'bear fruit', *wášte* 'be

good', *sápa* 'to be black', *tʰāka* 'to be large', *smáha* 'to be deep (as a valley)', *ūká* 'to be in a horizontal position', etc.

In this system, animate subject selection is a property of all members of the subject-inflecting predicate category and thus is a necessary condition on category membership, but it fails as a sufficient condition, since the object-inflecting category of the system also contains predicates which select animate subjects.

3.2 CHOCTAW SPLIT INTRANSITIVES. A similar structure of split intransitives is found in Choctaw (Muskogean), as analyzed in Davies 1986. Davies claims that members of the object-inflecting class of intransitive predicates all refer to events in which the single participant is nonvolitional; there is no selection restriction for animate subjects (Davies, p.c.). A sample of these predicates is given in (3).

3) CHOCTAW OBJECT-INFLECTING INTRANSITIVES (Davies 1984, 35)
    hohchafo 'hungry', *chaːha* 'tall', *abiːka* 'sick', *kapassa* 'cold',
    basha 'cut', *sipokni* 'old', *kobafa* 'break', *hottopa* 'hurt',
    laksha 'sweat', *albasha* 'suffer', *hohfaya* 'ashamed', *yoshoba* 'lost'

Members of the subject-inflecting class may be volitional or nonvolitional; it includes volitional predicates such as *ĩpa* 'eat', *washoha* 'play', *hilha* 'dance', *toksali* 'work', as well as nonvolitional predicates such as *illi* 'die', *ata* 'be/exist', most other existential predicates, and 'posture verbs' like *binili* 'sit, *hikiya* 'stand', and *itola* 'lie', among others. Just as in Dakota split intransitives, then, one class is semantically characterized but the other is not its complement, so that the property of selecting a nonvolitional participant is a necessary but not sufficient condition for membership in the object-inflecting category. This is a typical pattern among the split intransitive systems investigated in Merlan (1985); other systems which seem to be of this type include Seneca (Iroquoian), Arikara (Caddoan), and Eastern Pomo (Hokan).

4 INVERSION PREDICATES

4.1 SPANISH INVERSION PREDICATES. Examples parallel to these can be found in split transitive systems. A familiar example of this is so-called INVERSION predicates of Relational Grammar,[2] where a thematically salient noun phrase which is often characterized as an Experiencer is associated with a morphosyntax elsewhere associated with a Recipient in ditransitive clauses.[3] For example, Spanish has a class of predicates which encode experiencers with dative morphology in that they are manifested either as clitic pronouns from the dative series or as a noun phrase or tonic pronoun marked with the preposition *a* and crossreferenced by a dative clitic. A list of representative predicates from the inversion category in Spanish is given in (4).

4) SPANISH INVERSION PREDICATES

gustar 'to like', *parecer* 'to seem', *faltar* 'to be wanting, to lack', *hacer falta* 'to lack', *doler* 'to hurt', *encantar* 'to enchant, delight', *convenir* 'to agree, suit', *bastar* 'to be enough', *interesar* 'to interest', *ocurrir* 'to happen (to)', *importar* 'to be of concern', *impresionar* 'to impress, affect', *divertir* 'to entertain', *ofender* 'to offend', *sobrar* 'to have too much', *molestar* 'to bother', *disgustar* 'to disgust'

Spanish inversion predicates are restricted to expressions of physical or mental states; they can all be characterized as nonvolitional, and typically or exclusively selecting an animate subject in the role of Experiencer. However, we find predicates with the same semantic characterization in the class of predicates which do not have inversion morphosyntax. For example, many physical or mental states are expressed as statives with the copular verbs *ser* or *estar* and the Experiencer showing nominative morphosyntax and syntactic subject properties; some physical states are expressed with possessive morphosyntax using the verb *tener* 'to have', which again aligns the Experiencer with nominative morphosyntax and syntactic subject. Also, we find verbal predicates expressing physical or mental states with the Experiencer manifested as having nominative morphosyntax. These are illustrated in (5).

5) SPANISH PREDICATES WITH NOMINATIVE MORPHOSYNTAX

odiar 'to hate, dislike', *sentir* 'to feel', *amar* 'to love', *querer* 'to like, want, love', *temer* 'to fear, be afraid of', *creer* 'to believe', *dudar* 'to doubt', *esperar* 'to hope', *entender* 'to understand', *sufrir* 'to suffer', *estar contento* 'to be happy', *estar triste* 'to be sad', *ser loco* 'to be crazy', *ser inteligente* 'to be intelligent', *tener hambre* 'to be hungry', *tener fró* 'to be cold', *tener sueño* 'to be sleepy', etc.

The category associated with nominative morphosyntax and typified by the active transitive construction thus contains members with the defining properties of the inversion class, as well as other members without those defining properties, such as volitional predicates (*comer* 'to eat', *cantar* 'to sing', *tocar* 'to play an instrument', *escribir* 'to write', *escuchar* 'to listen', *comprar* 'to buy', *traer*, 'to bring', etc.), predicates which typically select or allow inanimate participants (*ser viejo* 'to be old', *crecer* 'to grow',

emitir 'to emit', *destruir* 'to destroy', etc.).   Thus, the properties of affectedness, nonvolition, and typical animacy of subject are necessary but not sufficient properties of the defined category of Spanish inversion predicates.

4.2 MALAYALAM INVERSION PREDICATES. There is also an inversion class of transitive predicates in Malayalam (Dravidian). It is characterized in Mohanan and Mohanan (to appear) as consisting of predicates with Experiencer subjects. A representative set of predicates from the inversion class is given in (6).

6) MALAYALAM INVERSION PREDICATES (Abrahams 1976)

iṣṭapeṭuka 'like', *eriyuka* 'experience hot taste', *kaaṛuka* 'experience itching sensation (throat)', *kaḻaykkuka* 'experience pain', *kuḻiruka* 'feel pain', *taḻaruka* 'be fatigued', *daahikkuka* 'experience thirst', *piṭikkuka* 'like', *boodhikkuka* 'approve', *rasikkuka* 'enjoy', *tooṇṇuka* 'feel', *naṣṭapeṭuka* 'lose', *nanassilaakkuka* 'understand', *labhikkuka* 'get', *okkuka* 'be able to', etc.

In their discussions of these predicates, however, Mohanan and Mohanan note that not all verbs having Experiencer subjects appear in the inversion class. Rather, a number of them appear in the category which has nominative morphosyntax and where the NP denoting the Experiencer shows subject properties. A few examples of such predicates are given in (7). This class also includes a range of other predicate-types, including predicates corresponding to 'serve', 'work', and 'give', along with other agentive predicates.

7) MALAYALAM AFFECTIVE PREDICATES IN ACTIVE TRANSITIVE CLASS
(Mohanan and Mohanan, 25; Abrahams 1976, 122)

talarṇṇu 'got tired', *santoosiccu* 'was happy', *ḍukkʰiccu* 'was sad', *bʰayappeṭṭu* 'feared/was afraid', *koopiccu* 'became angry', *ṇaṭunni* 'was shocked', *amparaṇṇu* 'perplexed'

Mohanan and Mohanan represent the difference between the predicates in (6) and those in (7) as a thematic structure difference, in that in the lexical representation of those in (6), the subject is Experiencer-Goal, while in (7), the subject is Experiencer-Theme. Given that both conceptualizations of the relation between an Experiencer and a state (state coming to Experiencer with Experiencer-Goal and Experiencer in state with Experiencer-Theme) are lexicalized in Malayalam, it isn't obvious that one can predict on the basis of the predicate itself which conceptualization will be lexicalized for which predicate. I would tentatively classify this predicate class system like the others examined here, in that it consists of a semantically specifiable class of predicates whose Experiencer subjects are associated with dative

morphosyntax and a neutral class of predicates whose subjects are associated with nominative morphosyntax. Other languages which appear to have a similar relation between a relatively small lexical category of semantically specifiable inversion predicates and a relatively large semantically heterogeneous category of predicates with nominative/accusative morphosyntax include Udi (Harris 1984a), and Choctaw (Davies 1986).

5 THE ARCHITECTURE OF DEFINED/NEUTRAL SYSTEMS. What these examples seem to show is that there may be a number of cases where the members of one category in a split predicate system can be said to share some semantic property or to have one property or more out of a set of properties which recur in their association with this set cross-linguistically, such as animacy, non-volitionality, affectedness, etc. In these systems, it seems that one category is semantically coherent in the sense that all members have a property or properties which they share with other set members, but the other category has both members which have and members which lack the defining property or properties of what I will call the DEFINED CATEGORY, and thus the second category is not semantically coherent. I'll call this category the NEUTRAL CATEGORY. A comparison of the relations between categories in a classical system, which I'll refer to as a DETERMINED system, and categories in a defined system is shown in (8) and (9).

8) STRUCTURE OF DETERMINED CATEGORY SYSTEMS

DETERMINED CATEGORY	COMPLEMENT CATEGORY
Necessary and sufficient conditions stated in terms of a set of properties	Necessary and sufficient conditions stated in terms of the absence of the properties of the determined category

9) STRUCTURE OF DEFINED CATEGORY SYSTEMS

DEFINED CATEGORY	NEUTRAL CATEGORY
Necessary conditions stated in terms of a set of properties	Members with and without the properties of the defined (=sufficient condition: absence of properties of the defined category)

6 FLUID PREDICATES. Another phenomenon often associated with split predicate systems and which sometimes presents a problem to Strong Semantic Alignment is the existence of a subset of predicates which can appear with the diagnostic morphosyntax of either category, with a concomitant semantic contrast.

6.1 MALAYALAM INVERSION CONSTRUCTIONS. In the case of Malayalam, there is a class of fluid predicates which appear in both the inversion and neutral patterns, as illustrated in (10-12).

10) a   *enikkə piṭiccu*   'I liked' (Abraham 1976, 133)
         1-dat   liked
     b   *ɴaan   piṭiccu*   'I held'
         1-nom   liked
11) a   *enikkə coṛiññu*   'I itched' (Abraham 1976, 133)
         1-dat   exper. itching
     b   *ɴaan   coṛiññu*   'I scratched'
         1-nom   exper. itching
12) a   *enikkə uraŋŋaam*   'I have permission to sleep'
         1-dat   sleep-MODALITY[4] (Mohanan 1982, 541)
     b   *ɴaan   uraŋŋaam*   'I promise to sleep'
         1-nom   sleep-MODALITY

The generalization proposed in Mohanan and Mohanan for a more restricted range of dative subject constructions[5] would seem to work predictively here, in that Experiencer-Goal is a plausible thematic characterization of the subject in the **a** sentences, while in the **b** sentences, some role other than Goal is associated with Experiencer, perhaps Effector. In Malayalam, then, the fluid predicates have an explicit semantic contrast while there is no contrast as such between the defined and neutral categories. The alignment between them is motivated: put simply, given the semantic contrast in fluid predicates of this type and given the semantics of the defined category, only one possible alignment between the semantically contrasting pairs of items and the polarity established by the defined category is consistent, as shown in (13).

13) MALAYALAM FLUID PREDICATE ALIGNMENT

category contrast:	Experiencer-Goal (DEFINED)	unspecified (NEUTRAL)
fluid predicate contrast:	Experiencer-Goal	Experiencer-Effector

6.2 CHOCTAW FLUID INTRANSITIVES. A fluid predicate pattern also exists in the split intransitive system of Choctaw, discussed above. Some predicates are fluid between subject-inflection and object-inflection, as illustrated by the examples in (14) and (15).

14) CHOCTAW FLUID PREDICATES (Davies 1984, 36)
     a   *sa - ttola -tok*   'I fell'
         1ACC fall PAST
     b   *ittola - li - tok*   'I fell (on purpose)'
         fall   1NOM PAST

15) a *sa – habishko – h*   'I sneezed'
       1ACC sneeze PRED
    b *habishko – li – h*   'I sneezed (on purpose)'
       sneeze 1NOM PRED

Here, the contrast is again between volitional and nonvolitional participation on the part of the subject. This is a more specific semantics than that of the defined and neutral categories, because although all members of the defined category of object-inflecting predicates share the property of selecting a nonvolitional subject, the neutral category is not uniquely characterized as selecting a volitional subject. Once more, however, in spite of the discorrelation between the category system architecture and the contrastive use of its associated morphosyntax, the relation between the categories and the fluid cases is motivated, in that the only consistent alignment would be that represented in (16).[6]

16) CHOCTAW FLUID PREDICATE ALIGNMENT

lexical category contrast:	volitional (DEFINED)	unspecified (NEUTRAL)
fluid predicate contrast:	volitional	nonvolitional

7 DAKOTA SECONDARY CATEGORY SYSTEMS. Dakota has no set of nonderived fluid predicates in its split intransitive system. However, there are forms derived by reduplication from a subset of the stems in the object-inflecting category (the neutral category in the Dakota system). These can take either the stress pattern typical of reduplicated object-inflecting verbs (second-duplicate stress) or of reduplicated subject-inflecting verbs (first-duplicate stress). Boas and Deloria 1941 list 18 pairs of this type; some representative examples of this are given in (17).

16) NEUTRAL CATEGORY STRESS   DEFINED CATEGORY STRESS
     a xopxópa   xópxopa
        'to be goodlooking'   'to pose, try to appear at one's best'
     b pispíza   píspiza
        'ability to whistle or squeal'   'he whistles or squeals'
     c snasná   snásna
        'possess a metallic, tinkling   'give off a metallic, tinkling
        quality'   sound'
     d snisní   snísni iyáya
        'to be cold to the touch'   'he turns cold'

e	snisníza	snísnis
	'to be in a collapsed state'	'gradually collapsing'
f	šnišníža	šníšniš áyapi
	'to be in a faded, wilted cond.'	'they are getting faded'

Merlan discusses this set of contrasts and claims that the derived forms show a volitional distinction, apparently on the basis of examples like (16)a and b. However, a closer examination of the set of forms given in Boas and Deloria show that many of the derived forms with the stress pattern typical of derivation from the defined category allow or select inanimate subjects, indicating that the contrast is actually between a state and a process (e.g. ability to squeal vs. emitting a squeal, being collapsed vs. becoming collapsed, etc.).

Pronominal prefixes of the Dakota possessive paradigm used for inalienable possession are also taken from the object-inflecting verbal paradigm (the neutral category), so that the second person possessive pronoun is invariably *ni-* . However, the body part lexicon divides into two classes with respect to the selection of the first person singular pronoun. The larger class (by a ratio of about 2:1) takes the first person object-inflecting form *ma-* , while the smaller class takes the prefix *mi-* which is not found in the verbal paradigm. Boas and Deloria (1941, 127) claim that the *mi-* prefix is used with body parts which are conceived of as 'particularly subject to willpower'. On this basis Merlan 1985 claims that the split in the body part lexicon is based on potential possessor control, with *mi-* used for body parts subject to control and *ma-* for all others. Thus, for example, *mit'ǎc'ǎ* 'my body' can be quite naturally viewed as controllable relative to *mahúhu* 'my bone' or *mawé* 'my blood'. However, if this system is viewed as a determined system, there are a number of apparent anomalies, as can be seen from the sample list in (17).

17) DAKOTA INALIENABLE POSSESSION (Boas and Deloria 1941, 128)

a	miísto	mač'éca
	'my arm'	'my leg'
b	minápe	mašášte
	'my hand'	'my little finger'
c	minǔẏe	map'ásu
	'my ear'	'my nose'
d	mič'ǎte	map'é
	'my heart'	'my liver'

We would have to assume that this categorization is heavily culture-specific so that, for example, some sensory organs such as 'ear' are viewed as more

controllable and others such as 'nose' are less so, that some internal organs such as 'heart' are viewed as more controllable and some such as 'liver' are less so, that 'arms' are more controllable than 'legs', that 'hands' are more controllable than 'fingers', etc. I don't deny the legitimacy of culture-specific criteria for category membership, but I think such an account would overstate the role of culture in this particular category system. I would claim, rather, that the only category whose membership needs to be accounted for is the *mi-* prefixing category and that the defining property is high perceived controllability. On the other hand, the *ma-* prefixing category needs no semantic or cultural explanation -- if it is viewed as a neutral category. Three body part terms are listed under both the *ma-* and the *mi-* prefixes. These are *miíte* 'my facial expression' vs. *maíte* 'my face', *misí/masí* 'my foot', and *mihó/mahó* 'my voice'; only the first of these is given a translation which clearly indicates the semantic distinction, but from the description, I will assume that the same kind of contrast is present in all three, and that that contrast is controllability vs. noncontrollability. The alignment between the lexical category system of Dakota split predicates, the semantic contrast of the derived fluid predicates, the body part possessives and the fluid body parts is given in (18).

18) <u>DAKOTA CATEGORY ALIGNMENT</u>

category distinction:	selects animate subject (DEFINED)	no subject selection restriction (NEUTRAL)
fluid predicate contrast:	process	state
inalienable possession:	controllable	no implication of control
fluid body parts:	controllable	noncontrollable

The semantic connection between the necessary property of the split intransitive defined category and the other secondary contrasts is somewhat less direct than in the cases of Choctaw and Malayalam examined above. Nonetheless, there is still a very natural relationship, if one assumes a typical connection between animacy and potential control and between potential control and process initiation. I don't think that any linguist would fail a matching test if asked to align the split predicate category system with the other systems. It is not surprising that the alignment relationship may be less direct in such a case, because the morphosyntactic relationship is also less direct, mediated either by derivational processes in the case of the fluid derived intransitives or by the contrasts of a distinct bound pronominal paradigm with only a partial form correspondence to markers of the split intransitive paradigm.

9 MOTIVATED ALIGNMENT. To summarize, while there may be some split predicate systems which are compatible with Strong Semantic Alignment, there seems to be a significant set of others which are not, but which nonetheless show a specifiable architecture.   We can still impose strong and plausible conditions on category systems, stated in terms of the architecture of DEFINED/NEUTRAL systems as represented in (9).   Although the systems explored here have been essentially binary, and the properties of the defined categories have been semantic, I will state Motivated Lexical Category Alignment somewhat more generally, as in (19).[7] (20) states a condition on the relationship between a lexical category system and its associated fluid system.

19) MOTIVATED LEXICAL CATEGORY ALIGNMENT
For a set of categories, all but one is defined. Defined categories are motivated in the sense that they are defined by a set of properties such that all members necessarily share one or more of those properties.
20) MOTIVATED FLUID ALIGNMENT
The relationship between defined and neutral categories in a lexical category system and the semantic contrast(s) in a corresponding fluid system is motivated.

If fluid elements are considered ambicategorial rather than members of an independent, unrestricted category, then the structural relation between the lexical category architecture and the fluid elements follows directly from the assumption that fluid elements, being ambicategorial, are members of the defined category and therefore share the necessary properties of that category. This provides a strong limiting condition on the relationship between lexical category systems and fluid predicate systems in that it anchors the fluid predicate contrast to the semantics of the defined category and leaves the contrasting category free to assume the complement value (though there is no reason that it must do so -- I leave this issue open).   The Dakota secondary contrasts based on the split intransitives do not violate this, since they represent separate and more distantly related systems. The relationship, then, between the DEFINED/NEUTRAL lexical system and its corresponding fluid systems seems to be that the fluid systems more closely approximate the structure of a DETERMINED/COMPLEMENT system.

I hope to have shown here that there is a significant and highly constrained external architecture to the categories in lexical category systems which are not determined systems.   This aspect of the category architecture of the lexicon is not autonomous but can only be fully understood when an account of this architecture is integrated into a more comprehensive account of the use to which these category distinctions are put in the expression of meaning.   An examination of fluid elements and secondary contrast systems gives insight into how these categories function as a support structure for semantic contrast.

303

NOTES

*I am grateful to Bill Davies for clarifying aspects of his analysis of Choctaw for me and to K.P. Mohanan for an enlightening discussion of Malayalam and the nature of predictability; all interpretations are my own. I want to thank Marshall Lewis, whose careful reading of a draft of this paper and helpful discussions on this topic have led to significant improvements.

[1] I won't be concerned with the internal structure of categories, although I will make the assumption that categories do have significant internal structure of the type examined by George Lakoff (1987) and others.

[2] I don't attempt here to distinguish between systems where the inverted NP has significant subject properties and where it does not.

[3] Inversion predicates cross-linguistically belong to one or more related semantic types (Harris 1984b, Sridhar 1976), including physical states: be cold, be tired, be thirsty, etc.; mental states: be amused, like, hate, understand, believe, etc.; modality: necessity, potential, etc., possesion, and existence.

[4] Note that in (14) it is not the predicate meaning 'sleep' which is fluid; rather, it is the modality suffix -aam which is the fluid predicate, and it is the contrastive modalities expressed by the English glosses 'permission' and 'promise' which show the semantic pattern of contrast.

[5] K.P. Mohanan has informed me (p.c.) that a more comprehensive account of dative constructions based on the preliminary semantic analysis of Mohanan and Mohanan (to appear) is in their future research program.

[6] In at least some split predicate systems, the specific semantics of the fluid predicate contrast may also be motived in part by the internal structure of the neutral category. In all cases that I know of which involve split transitive systems with inversion, the neutral category contains predicates denoting all or most prototypical transitive event types (along the lines specified in Hopper and Thompson 1980, for example). Under the assumption that the prototypical transitive event type involves a volitional actor, it isn't surprising to find an alignment between the neutral category and volitionality.

[7] A preliminary exploration of a wider range of lexical categories in predicate systems as well as a wider range of structures in the languages examined here seems to indicate that the more general statement referring to category systems more complex than binary systems is warranted. The assumption that form similarity is one of the principles of internal category organization justifies the lack of restriction to semantically defined properties.

REFERENCES

Abrahams, P.T. 1976. Pseudo-datives in Malayalam. In *Dravidian Case System*, S. Agesthialingom and K. Kushalappa Gowda, eds., 121-136. Annamalaingar: Annamalai University.

Boas, Franz and E. Deloria. 1941. *Dakota Grammar* . Memoirs of the National Academy of Sciences, Vol. 23, Second Memoir. Washington, D.C.

Chun, Soon Ae, and David A. Zubin. This Volume. Experiential vs. agentive constructions in Korean narrative.

Davies, William. 1986. *Choctaw Verb Agreement and Universal Grammar.* Dordrecht: D. Reidel.

Dixon, R.M.W. 1979. Ergativity. *Language* 55:59-138.

Durie, Mark. 1987. Grammatical relations in Acehnese. *Studies in Language* 11:365-399.

Grimshaw, Jane. 1989. *Argument Structure* . Brandeis University ms.

Harris, Alice. 1984a. Case marking, verb agreement, and inversion in Udi. In Perlmutter and Rosen, eds., pp. 243-258.

_____. 1984b. Inversion as a rule of universal grammar: Georgian evidence. In Perlmutter and Rosen, eds., pp. 259-291.

Hopper, P. and S. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56:251-299.

Lakoff, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind* . Chicago: University of Chicago Press.

Merlan, Francesca. 1985. Split intransitivity. In J. Nichols and A. Woodbury, eds., *Grammar Inside and Outside the Clause* , pp. 324-362. Cambridge: Cambridge University Press.

Mohanan, K.P. 1982. Grammatical relations and clause structure in Malayalam. In Joan Bresnan and R. Kaplan, eds., *The Mental Representation of Grammatical Relations* , pp. 504-589. Cambridge, Mass.: The MIT Press.

_____ and Tara Mohanan. To appear. Dative subjects in Malayalam: Semantic information in syntax. In K.P. Mohanan and M. Verma, eds., *Experiencer Subject in South Asian Languages* , CSLI.

Perlmutter, David and Paul Postal. 1984. The 1-advancement exclusiveness law. In Perlmutter and Rosen, eds., pp. 81-125.

_____ and Carol Rosen, eds. 1984. *Studies in Relational Grammar 2* . Chicago: University of Chicago Press.

Rosen, Carol. 1984. The interface between semantic roles and initial grammatical relations. In Perlmutter and Rosen, eds., pp. 38-77.

Schwartz, Linda. To appear. On the syntactic and semantic alignment of attributive and identificational constructions. In *Advances in Role and Reference Grammar* , R. Van Valin, Jr., ed. Amsterdam: John Benjamins.

Sridhar, S.N. 1976. Dative subjects. In *Papers from the Twelfth Regional Meeting of the Chicago Linguistic Society*, 582-593. Chicago: University of Chicago Press.

Van Valin, Jr., Robert. To appear. Semantic parameters in split intransitivity. *Language.*

IS THERE OBJECT-TO-SUBJECT RAISING IN CHINESE?

Dingxu Shi
University of Southern California

The comparison of similar linguistic phenomena across the world's languages has always been a resourseful means for the study of grammar. There is, however, no a priori guarantee that what appear to be similar are indeed sufficiently correspondent to justify the crosslinguistic comparison of structure. Sometimes it will turn out that apparently comparable phenomena are, under scrutiny, rather different in nature. The seeming resemblance between the raising structure in English and the alleged object-to-subject raising structure in Chinese is such a case.

1. The Object-to-Subject Raising Analysis

The process of object-to-subject raising is also known as tough-movement. It is usually associated with the standard theory of generative grammar, but it has attracted much attention in current linguistic studies with regard to language universals and the issue of empty categories in formal syntax (e.g. Comrie and Mathews 1989, Lasnik and Uriagereka 1988). The raising analysis is prompted by the fact that in certain English sentences, such as the one in (1a), the surface matrix subject does not bear any semantic relation to the matrix verb. As a matter of fact, the collocation of the surface subject and matrix verb is usually nonsensical or has a different interpretation, as in the case of (2).

(1) a. John is easy to please.

b. It is easy to please John.

(2) John is easy.

In terms of thematic relation, the surface subject in such a sentence is apparently the patient NP of the embedded verb. Sentence (1a) seems to have the same truth value as (1b) does. In the early days of generative tranformational grammar, this generalization is captured by postulating the surface subject as the embedded object in deep structure. It is then moved to the surface subject position by an object-to-subject raising rule (Postal 1974). The movement is treated in a different manner in the current government-binding theory, but the crucial assumption remains the same, namely, the surface matrix subject is related to the embedded object position (Chomsky 1981, 1986).

What have inspired the object-to-subject raising analysis in Chinese are sentence pairs like (3) (from Li 1985) and (4), which are apparently parallel to the English pair in (1). The two sentences in such pairs seem to have the same truth value, and they have exactly the same constituents. The only difference is the order of these constituents. The (a) sentence in these pairs consists of a sentential subject and an adjectival predicate. In the corresponding (b) sentence, the verb of the subject clause

in (a) appears in the sentence final position. The adjectival predicate of (a) now occurs between the initial NP and the final verb of (b).

(3) a. [Chongfu   zheige   gushi] hen   nan.
       repeat   this Cl.   story very   difficult
       'To repeat this story is very difficult.'

    b. Zheige   gushi hen   nan   chongfu.
       this Cl.   story very   difficult   repeat
       'This story is very difficult to repeat.'

(4) a. [Anshi   wancheng   jihua]   bu   rongyi.
       on time   complete   project   not   easy
       'To complete the project on time is not easy.'

    b. Jihua   bu   rongyi anshi   wancheng.
       project   not   easy   on time   complete
       'The project is not easy to complete on time.'

According to the object-to-subject raising analysis, the sentence initial NP in the (b) sentence has become the surface subject, while the adjective phrase after it remains unchanged as the matrix predicate. In other words, the subject clause in the (a) sentence is now the embedded clause. This movement is triggered by the so-called raising verbs, namely, the adjective in the matrix predicate of the (a) sentences, which include keneng 'possible', hen nan 'very difficult' and rongyi 'easy'. How the movement is carried out differs from analysis to analysis. In Hou (1979), the (b) sentence is generated in the same deep structure as (a) is, i.e., with a subject clause. The object-to-subject raising rule postposes the VP of the subject clause and creates an embedded clause. In some other analysis, (e.g. Li 1985), the (b) sentence is generated with an embedded clause and an empty matrix subject position. The object of the embedded clause is then raised to become the matrix subject. If the embedded object is not raised, the whole embedded clause will surface as the subject.

As part of the parallel analysis, it has been claimed that the raising verbs, i.e., the adjectival predicates, only take nonfinite complement clause (Li 1985, Goodall 1983). Ungrammatical sentences like (5) and (6) (from Li 1985:129) are cited as evidence for this claim. The argument is that a nonfinite verb does not inflect. The unacceptable status of (5) is therefore caused by the aspect marker le attached to the embedded verb. It is a standard assumption of the government-binding theory that the subject position of a nonfinite clause cannot be filled by a lexical NP, because the NP there cannot get case and will not pass the Case Filter. The sentence in (6) is thus ungrammatical since the subject position of the embedded clause is filled with an NP.

(5) *Zheige   gushi   bu   rongyi chongfule.
    this Cl.   story   not   easy   repeat Asp.
    'This story is not easy to have repeated.'

(6) *Zheige   gushi   bu   rongyi [xuesheng   chongfu].
    this Cl.   story   not   easy   student   repeat
    'This story is not easy for the students to repeat.'

2. Problems Of The Raising Analysis

The parallel between the English raising structure and its alleged Chinese counterpart does not go any further. Upon careful examination, the Chinese raising structure turns out to have many unique properties of its own. A prominent one is that the one-to-one relationship between the two sentences in pairs like (3) is not always available. The sentence in (7a), for example, is related to the sentence in (4a) and the two have almost identical structure. There is, however, no counterpart for (7a) with a raised object. Neither (7b) nor (7c) is acceptable. The object NP of the subject clause is not inside any island. The object-to-subject raising rule simply fails to apply in this case.

(7) a. [Neng   wancheng   jihua]   bu   rongyi.
       can   complete   project   not   easy
       'To be able to complete the project on time is not easy.'

    b. *Jihua   neng bu   rongyi   anshi   wancheng.
       project   can   not   easy   on time   complete

    c. *Jihua   bu   rongyi neng anshi   wancheng.
       project   not   easy   can   on time   complete

Another unusual property of the alleged raising process in Chinese is that it seems to be able to raise an object across clause boundaries. In sentence (8a), the NP zheige jihua 'this project' is the object of an embedded clause inside the subject clause. In the corresponding (8b), the same NP occurs in the sentence initial position. Under the rasing analysis, this NP would be considered as being raised from the embedded object position to the matrix subject position. This kind of 'super-raising' is rather peculiar in the sense that it is not allowed in English and other languages. The restriction on raising across clause boundaries is usually given as the explanation for why the English sentence in (9b) cannot be derived from (9a) (cf. Postal 1974). The object-to-subject raising in Chinese seems to differ from that of English in this respect.

(8) a. [Lijie   tamen weishenmo   bu   anshi
       understand   they   why   not   on time
        wancheng   zheige   jihua]   bu   nan.
       complete   this Cl.   project   not   difficult
       'It is not difficult to understand why they do not
        complete this project on time.'

b.

Zheige	jihua	bu	nan	lijie
this Cl.	project	not	difficult	understand
tamen	weishenmo	bu	anshi	wancheng.
they	why	not	on time	complete

'As for this project, (one) has little difficulty in understanding why they do no want to complete (it) on time.'

(9) a. It is easy to see that John loves Mary.

b. *Mary is easy to see that John loves.

The third unexpected property of the object-to-subject raising structure in Chinese is that the raising seems to be optional. Notice that the sentence initial NP in the (b) sentence of (3), (4) and (8) can always occur after the sentence final verb without changing the acceptability and interpretation of the sentence, as shown in (3c), (4c) and (8c) below. The NP in question is apparently not obliged to raise even when the sentence initial position is not occupied by any NP. This is in contrast with the English object-to-subject raising, which is obligatory in the sense that the matrix subject position has to be filled, otherwise the sentence will not be acceptable, as in the case of (1c).

(1) c.* Is easy to please John.

(3) c.

Hen	nan	chongfu	zheige	gushi.
very	difficult	repeat	this Cl.	story

'(It is) very difficult to repeat this story.'

(4) c.

Bu	rongyi	anshi	wancheng	jihua.
not	easy	on time	complete	project

'(It is) not easy to complete the project on time.'

(8) c.

Bu	nan	lijie	tamen	weishenmo
not	difficult	understand	they	why
bu	anshi	wancheng	zheige	jihua.
not	on time	complete	this Cl.	plan

'(It is) not difficult to understand why they do not complete this project on time.'

The most extraordinary feature of the alleged object-to-subject raising in Chinese is the dubious status of the raised object. According to the raising analysis, the sentence initial NP in (3b) and (4b) is the surface subject, even though it has the thematical role of patient of the embedded verb. A problem arises when the agent NP of the same verb is brought into consideration. The agent NP is not present in the data examined so far, but its absence is by no means obligatory. In a sentence like (3b), the agent NP of the embedded verb can occur legitimately in a position preceding the raising verb, either before the patient NP or after it, as in the case of

(3d) and (3e) respectively. The position of the patient NP is not a factor here. When it appears after the verb, the agent NP can still be present in a preverbal position, as in (3f) below.

(3) d. <u>Wo</u>   <u>zheige</u>   <u>gushi</u> <u>hen</u>   <u>nan</u>   <u>chongfu.</u>
     I   this Cl.   story very difficult   repeat
     'I, this story, can hardly repeat.'

   e. <u>Zheige</u>   <u>gushi</u> <u>wo</u>   <u>hen</u>   <u>nan</u>   <u>chongfu.</u>
     this Cl.   story I   very difficult   repeat
     'This story, I can hardly repeat.'

   f. <u>Wo</u>   <u>hen</u>   <u>nan</u>   <u>chongfu</u>   <u>zheige</u>   <u>gushi.</u>
     I   very difficult   repeat   this Cl.   story
     'I can hardly repeat this story.'

It is generally agreed that when the agent NP and the patient NP of an active verb are both present in the same sentence, the agent NP, but not the patient NP, will assume the grammatical function of subject (Fillmore 1968, Keenan 1976). There is no obvious reason to consider (3d), (3e) and (3f) as exceptional cases. The contrast between (10a) and (10b) provides some additional evidence for the argument that the raised object in such sentences does not behave like a subject.

(10) a.*[<u>Zheige</u>   <u>xiaotou</u>]$_j$ <u>hen</u>   <u>nan</u>
      this Cl.   thief   very difficult
     <u>zai</u> <u>ziji</u>$_j$ <u>jia</u>   <u>zhuadao.</u>
     at   self   home catch
     'This thief is difficult to catch at his own home.'

   b. [<u>Zheige</u>   <u>xiaotou</u>]$_j$ <u>jingcha</u>$_i$ <u>hen</u>   <u>nan</u>
      this Cl.   thief   police   very difficult
     <u>zai</u> <u>ziji</u>$_{*j/i}$ <u>jia</u>   <u>zhuadao.</u>
     at self   home catch
     'This thief, the policeman$_i$ can hardly
     catch him at his$_i$ own home.'

Inside the locative prepositional phrase of (10a) and (10b), there is a reflexive <u>ziji</u> 'self'. The indexing and interpretation of (10b) shows that the anaphor can take as its antecedent the agent NP <u>jingcha</u> 'policeman', but not the patient NP <u>xiaotou</u> 'thief'. Even when the agent NP is absent, the sentence initial patient NP still cannot function as the antecedent of the anaphor, as shown by the unacceptable status of (10a). It is a well known fact that the anaphor <u>ziji</u> 'self' takes only a c-commanding subject as its antecedent (Tang 1988, Cole et al. 1990). The pattern in (10) indicates clearly that the agent NP is the subject, but the preverbal patient NP is not.

If the argument is correct that the preverbal patient NP in sentences like (3b) does not have the status of a subject, the object-to-subject raising analysis in Chinese will lose most of its validity. An alternative account has to be found for the sentence initial patient NP.

3. An Alternative Analysis

Recall that the one-to-one relation is not always available between the two sentences in pairs like (3a) and (3b), but a one-to-one relation can always be established between the two sentences in pairs like (3b) and (3c) (repeated below). It is possible that (3b) is syntactically related to (3c), and the apparent correspondence between (3a) and (3b) is just accidental. Given this observation, an alternative analysis is available.

(3) a. [Chongfu   zheige   gushi] hen   nan.
       repeat   this Cl.   story   very   difficult
       'To repeat this story is difficult.

   b. Zheige   gushi   hen   an   chongfu.
      this Cl.   story   very   difficult   repeat
      'This story is difficult to repeat.'

   c. Hen   nan   chongfu   zheige   gushi.
      very   difficult   repeat   this Cl.   story
      '(It is) very difficult to repeat this story.'

It is generally agreed that the unmarked position for the direct object of a Chinese transitive verb is after that verb, at least in the surface structure. The patient NP in (3c) is therefore in its basic position, and the one in (3b) is in a derived position. The proposal of this paper is that the patient NP in (3b) is in a topic position.

This alternative analysis provides a straightforward account for the phenomena discussed so far in this paper. Topicalization in Chinese is an optional process in that no NP will be forced to move to the sentence initial position on pure syntactic ground. It is therefore legitimate for a patient NP to either stay in its normal postverbal position or move to the sentence initial topic position, as in (3b) and (3c) respectively.

Under the topicalization analysis, the problematic super-raising in (8b) becomes a predicted result. Topicalization in Chinese is an unbounded process that can move an NP across boundaries of embedded clauses as long as no island violation is involved. It is therefore quite natural for the object of the embedded clause to move to the sentence initial topic position. The only adjustment is that (8b) is no longer analysed as syntactically related to (8a). It is now considered as a counterpart of the sentence in (8c).

(8) b. <u>Zheige</u>   jihua   bu   <u>nan</u>   <u>lijie</u>
      this Cl.   plan   not   difficult   understand
      <u>tamen</u>   <u>weishenmo</u>   bu   anshi   <u>wancheng.</u>
      they   why   not   on time   complete
      'As for this project, (one) has little difficulty to understand
      why they do not complete (it) on time.'

   c. <u>Bu</u>   <u>nan</u>   <u>lijie</u>   tamen weishenmo
      not   difficult   understand   they   why
      <u>bu</u>   anshi   <u>wancheng</u>   zheige   <u>jihua.</u>
      not   on time   complete   this Cl.   project
      '(One) has little difficulty to understand why they
      do not complete the project on time.'

The simultaneous presence of the raised object and its agent NP has been the most serious challenge to the object-to-subject raising analysis. It poses no threat at all to the approach of this paper. The topic is usually considered as occupying a position outside the comment clause (S or IP), but not any position within the clause (Li and Thompson 1981, Chomsky 1981). When the patient NP of an active verb appears in the topic position, the subject position of the same verb will always be available for the agent NP to occur. It is quite common for a topicalized patient NP and its agent NP to cooccur, so that either (3d) or (3e) is normal.

Two issues have to be addressed here. One is the structure of sentences like (3d) (repeated below), where the patient NP occurs between the verb and the agent NP without any overt marker. This construction is rather common in Chinese, but is not well understood yet. A possible analysis is to treat the patient NP as a secondary topic (cf. Tsao 1977) or a focus NP. The agent NP will remain as the subject of the clause. The sentence will have a structural representation in the shape of (11a). Another possibility is to consider the patient NP as being topicalized first, and the agent NP being topicalized after it. This process will produce the structure of (11b). Either analysis is compatible with the approach in this paper.

(3) d. <u>Wo</u>   zheige   <u>gushi</u>   hen   <u>nan</u>   <u>chongfu.</u>
      I   this Cl.   story   very   difficult   repeat
      'I, this story, can hardly repeat.'

(11) a. NP_{subj} [$_{VP}$ $NP_{topic/focus}$ [$_{VP}$ V t]]
    b.   [$_{S'}$ NP_{topic1} [$_{S'}$ NP_{topic2} [$_{S}$ t_1 + V + t_2]]]

The other issue is related to the agent NP in sentences like (3b) and (3c). Under the raising analysis, the so-called raising verb is the matrix verb in (3b), and the verb that subcategorizes for the patient NP is nonfinite so that it cannot take a lexical subject. The claim follows that the agent NP must not appear in the sentence. It has been shown in the previous section that the claim is counterfactual and the agent NP can always appear in the same sentence with the patient NP. The topicalization analysis of this paper has the advantage of not excluding the presence of agent NP,

but it does not provide any direct explanation for the phenomenon that the agent NP can be absent and that it cannot occur between the raising verb and the embedded verb in (3b).

The account for these facts is supplied by some independently motivated analyses. It has long been noticed that an argument in a Chinese sentence can always be realized as a zero form if it is related to an entity being discussed in the discourse, namely, the so-called discourse topic. When the zero form occurs in the subject position, it can also stand for a universal quantifier, i.e., the NP <u>renheren</u> 'anybody' (Tsao 1977, Li and Thompson 1981, Huang 1984). In other words, the subject position of any Chinese sentence can be filled with a zero form, and sentences like (3b) are only a subtype of empty subject clauses. They differ from the other zero subject sentences in that their object NP has been topicalized, but that distinction bears no consequence on the occurrence of zero subject.

As for the claim that no NP can occur between the raising verb and the embedded verb in sentences like (3b), the explanation is provided by the alleged raising verbs themselves. Notice that the claim is not an accurate generalization. Although an agent NP cannot occur between certain raising verbs and the embedded verb, it can do so in cases like (12b). <u>Keneng</u> 'possible' is said to be a typical raising verb (Hou 1979), but it obviously allows the agent NP of the alleged embedded verb to occur after it. Apparently, there are two types of raising verbs with regard to the position of the agent NP in question. This classification coincides with another well known distinction. It has been pointed out that what used to be called preverbal adverbs consist of two types, adverbs like <u>keneng</u> 'possibly' and modals like <u>rongyi</u> 'easy' (Lü et al. 1983). A Chinese modal can only occur between the subject and the VP of its own clause, while an adverb can usually occur either between the subject and the VP or in a sentence initial position.

(12) a. <u>Zheige</u>   <u>gushi</u>   <u>keneng</u>   <u>tingguole.</u>
      this Cl.   story   possible   hear Asp.
      'This story (one) possibly has heard.'

   b. <u>Zheige</u>   <u>gushi</u>   <u>keneng</u>   <u>ni</u>   <u>tingguole.</u>
      this Cl.   story   possible   you   hear Asp.
      'This story, possibly you have heard.'

   c. <u>Zheige</u>   <u>gushi</u>   <u>ni</u>   <u>keneng</u>   <u>tingguo.</u>
      this Cl.   story   you   possible   hear Asp.
      'This story, you possibly have heard.'

Given the topicalization analysis, the so-called raising verbs do not have to be treated as matrix verbs. They can be considered as adverbs and modals. The adverbs allow the subject NP to precede or follow it, as in (12), but the modals prohibit the subject from occurring after it.

313

Incidently, the glossary usually given to sentences like (3b) is misleading. The sentence in (3b) should not be translated as 'The story is very difficult to repeat', but should be 'The story, (one) can hardly repeat it.'

4. Conclusion

The main argument in this paper is that the alleged object-to-subject raising in Chinese is not comparable to its English counterpart with respect to structure. It is a case of topicalization.

With respect to function, however, there may be some similarity between the object-to-subject raising in English and the topicalization in Chinese. Topicalization is a means to emphasize a certain NP, by moving it to the sentence initial position. In a sense, the function of object-to-subject raising is also to emphasize an object NP, by raising it from an embedded position to the matrix subject position (Chafe 1976).

The process of topicalization is more general in the sense that it is not restricted to the movement of a particular NP. It could be speculated, then, that object-to-subject raising does not exist in Chinese because a more general process is always available to carry out the same function.

REFERENCES:

Chafe, Wallace. 1976. Giveness, contrastiveness, definiteness, subjects, topics, and point of view. In Subject and topics, ed. by Charles Li. New York: Academic Press, Inc.

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris Publications.

---- 1986. Barriers. Cambridge: MIT Press.

Cole, Peter, Gabriella Hermon and Li-May Sung. 1990. Principles and parameters of long-distanc reflexives. Linguistic Inquiry 21, 1-22.

Comrie, Bernard and Stephen Mathews. 1989. Prolegomena to a typology of tough movement. To appear in Festschrift for Joseph Greenberg, ed. by William Croft, Amsterdam: Benjamins.

Fillmore, Charles J. 1968. The case for case. In Universals in linguistic theory, ed. by Emmon Bach and Robbert T. Harms, 1-88. New York: Holt, Rinehard and Winston.

Goodall, Grant. 1983. A three-dimension analysis of coordination. Chicago Linguistic Society, 19.

Hou, John. 1979. Grammatical relations in Chinese. Ph.D. dissertation, University of Southern California.

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. Linguistic Inquiry 15, 531-574.

Keenan, Edward. 1976. Towards a universal definition of 'subject'. In Subject and topic, ed. by Charles Li. New York: Academic Press, Inc.

Lasnik, Howard and Juan Uriagereka. 1988. A course in GB syntax. Cambridge: MIT Press.

Li, Charles and Sandra Thompson. 1981. Mandarin Chinese: a functional reference grammar. Berkeley: University of California Press.

Li, M-D. 1985. Reduction and anaphoric relations in Chinese. Ph.D. dissertation, University of California, San Diego.

Shuxiang, Linding Li, Jian Liu, Jiyan Fan, Youwei Shi, Fangliang Fan, Zong Meng, Shujun Ma, Zhu Li, Jianmin Chen, Kaidi Zhan, Huaide Zheng and Baoxiang Tao. 1983. Hanyu babai ci (Eight Hundred Chinese words). Beijing: The Commercial Press.

Postal, Paul M. 1974. On raising: one rule of English grammar and its theoretical implications. Cambridge: MIT Press.

Tang, C-C. Jane. 1989. Chinese reflexives. Natural Language and Linguistic Theory 7, 93-121.

Tsao, Feng-fu. 1977. A functional study of topic in Chinese: the first step towards discourse analysis. Ph.D. dissertation, University of Southern California.

The Role of Particles in Japanese Gossip*
Ryoko Suzuki
University of California, Santa Barbara

I. Introduction
    In Japanese conversational interaction, participants use various sentence particles (Uyeno 1971:50, Clancy 1985:427, Oishi 1985:7) which index the speaker's affective and epistemological stance toward the statement he or she is making. These particles generally occur in utterance-final position, and convey additional information about the speaker's attitude toward what he or she is saying (Martin 1975:914, Oishi 1985:7). Although there have been some descriptions of the meanings of these particles (e.g. Uyeno 1971, Martin 1975, Tsuchihashi 1983), there is very little information on how they are actually used by speakers in interaction (for exceptions, see Oishi 1985, Seki 1987, Cook 1988). The purpose of this paper is to present the results of a conversational analysis showing how these particles are used by two interlocutors to influence each other's stance about a particular topic as well as to maintain solidarity in the course of the interaction. I will focus on one type of conversational activity, gossip, in Japanese.

II. The particles
    The particles to be discussed are **yo, sa, ne** and **na**. Based on previous descriptions of these particles using intuitive and natural data, I came up with a distillation of the meaning of each of these four particles: **Yo** gives the flavor of 'authority' to the speaker about the statement he is making; **sa** is an 'insisting' particle, with which the speaker tries to convince the listener of the claim he is making; **ne** seeks the listener's 'agreement'; and **na** has the effect of 'contrasting' one's statement without imposing it on the listener.
    These four particles are very commonly used in Japanese gossip discourse, defined here as a series of assessments about an absent third party. A number of interesting questions arise: When interlocutors' stances toward the absent third party are different, do they try to influence each other's stance? Do they also try to negotiate the tension between them? And if they do, what role do the particles play in negotiating stance differences in gossip?
    The data are a twenty-minute, audio-taped conversation between Hiroshi, a Japanese male graduate student, and Sachiko, a Japanese female graduate student, both studying in the U.S. Hiroshi and Sachiko are good friends. The conversation takes place in the living room of Sachiko's apartment in Los Angeles. They are talking about another Japanese male student, Takao, who is also studying in the States and Takao's attitude towards his female married classmate, Mariko, who has approached him by bringing him food every day. Mariko's husband has read her diary and found out that she likes Takao. So the husband has accused Takao of encouraging the relationship and has put pressure on Takao through the school administration to quit graduate school. In the data, Sachiko starts telling Hiroshi the story of what happened to Takao. Hiroshi and Sachiko have different stances

toward Takao: Hiroshi criticizes Takao for not firmly rejecting Mariko, the female classmate. Sachiko does not criticize Takao at all at the beginning, but comes closer to Hiroshi's stance as the conversation proceeds.

The analysis shows that the particles strongly interact with the functions of the statements in the discourse.

(1) **Yo**, the particle of 'authority'

Sachiko uses **yo** mostly when she tells Hiroshi the story she heard from Takao, as we see in Example 1:

#1
Sachiko: Demo kurasumeeto de mainichi kao awaserunda tte **yo**.
　　　　　'But (they are) classmates and see each other every day, (he told me) **yo**.'

By adding the authority particle at the end of the information she directly heard from the central figure of the gossip, Takao, Sachiko actually presents herself as an authority on this information, and uses the information to support her stance that Takao is not to blame. On the other hand, Hiroshi is a recipient of the information, and he uses **yo** when he makes critical assessments of Takao's behavior. Hiroshi thereby presents himself as an authority on the morality of what Takao is doing, as in Example 2:

#2
Hiroshi: Sore wa sekinin-nogare dekinai to omou **yo**.
　　　　　'(I) think (he) cannot avoid taking responsibility **yo**.'

It has been pointed out that **yo** presupposes that the speaker assumes the hearer does not know the information (Clancy 1985:430, Cook 1988:130). Oishi (1985:177) and Cook (1988:151), through the analysis of natural conversation, found that **yo** is used by a specialist to exert power over an ordinary person who does not have the knowledge in question. The use of **yo** in the present gossip data confirms their findings: Sachiko tries to exert power by using information she heard from Takao that Hiroshi does not know. Cook (1988:140), who observed mother-child conversation, found that **yo** can be used when the speaker points out a wrong assumption held by the addressee. Hiroshi's use of **yo** shows that he tries to be authoritative by attaching **yo** to his stance and further, pointing out that Sachiko is wrong.

(2) **Sa**, the 'insisting' particle

Sa is very frequently used by Hiroshi. **Sa**, as well as **ne**, is used not only at the end of the utterance, but also right after an NP or conjunction in the data:

#3
Hiroshi: Soo iu rifujin na te o tsukau tte iu no wa **sa**,
　　　　　'That (the husband) did such an unreasonable thing **sa**,'
　　　　　(attention **sa**)
　　　　datte sonoo gakkoo o tooshite **saa**,

'going through the school (administration) **saa**,'
(<u>attention</u> **sa**)

koo atsuryoku o kakeru tte iu no wa **sa**,
'putting pressure (on Takao) **sa**,'
(<u>attention</u> **sa**)

sonna no moo juubun me ni mieteru ja nai.
'it is self-evident that such a thing could happen.'

Nihon tte soo iu shakai nandashi **saa**.
'Japan is that kind of society (as you know) **saa**.'
(<u>insisting</u> **sa**)

Sa, rather than seeking agreement from the listener, has a different type of effect when used utterance- internally. The first three uses of **sa** in Example 3 are inserted in the middle of the statement. Hiroshi has not finished making his point yet. **Sa**, the 'insisting' particle, when used repeatedly in this way, signals 'keep paying attention to the following.' The fourth **sa** is attached to statement of a generalization. Hiroshi warrants his stance by making a clear generalization, and furthermore, pushes this generalization by adding **sa**, and generates the flavor of the 'obviousness' of his statements.

(3) **Ne**, the 'agreement' particle

Ne is the most frequent particle used by Sachiko, and the second most frequent particle used by Hiroshi. **Ne** is used not only utterance-finally, but can occur utterance-internally, for example, right after an NP, adverbial or conjunction. **Ne** functions on two levels. First on the level of content, **ne** occurs when the speaker shows agreement with, or acceptance of, what the other speaker says. **Ne** is also used when the speaker seeks the listener's agreement with what he is saying. Second, on an interactive level, the speaker uses **ne** to seek the listener's agreement on his holding the floor (Cook 1988:170, Clancy p.c.). **Ne** inserted in utterance-internal positions typically serves this function. Example 4 shows the use of **ne** on both content and interactive levels:

#4
Sachiko: Dakara kondo wa ojisan o tsuujite puresshaa o kakete kite,
'So this time (the husband) put pressure on (Takao) through (the wife's) uncle,'

nanka adomission no hoo kara **ne**,
'well the admission office **ne**,'
(<u>interactive</u> **ne**)

Takao-san wa kore ijoo tsuzukerarenai yoo ni tte iu puresshaa o **ne**,
'put pressure on Takao so that he cannot continue (his studies) anymore **ne**,'
(<u>interactive</u> **ne**)

kakete kitanda tte.
'(he said.)'

Hiroshi: Soryaa shooganai daroo **ne**.

'(I) guess it cannot be helped **ne**.' (<u>content</u> **ne**)

The first and second **ne** used by Sachiko utterance-internally are seeking agreement with Sachiko's holding the floor as well as seeking the listener's agreement with Sachiko's stance that Takao is a victim. Hiroshi reveals his stance by also using **ne**, seeking agreement from Sachiko.

(4) **Na**, the 'contrasting' particle
    Na seems to serve an interesting function in gossip discourse in which the participants' stances are constantly projected. **Na**, the 'contrasting' particle, is less imposing than **sa** or **yo**. However, similar to the primary stress on <u>I</u> in English, **na** in gossip highlights the contrast between the speaker's stance and the listener's. Example 5 contains both the <u>attention</u> **sa** (as in Example #3) and the 'contrasting' **na**:

#5
Hiroshi: Ma sorya datte **saa**.
           'Well, however **saa**.'
         Sonoo.
           '(How shall I say it.)'
         Jibun no- jibun ni ii zairyoo shika hanasanai deshoo.
           '(He) only says things advantageous to him- himself (don't you think).'

         Ano hito tte **saa**.
           'That guy (=Takao) **saa**.'
         Kyakkanteki ni mitara,
           'From the third party's point of view,'
         yappari soryaa are da to omou **naa**, aa.
           '(I) think it is evident **naa**.'
Sachiko: Suki ga atta tte?
           '(You mean Takao) allowed (Mariko) to approach (him)?'
Hiroshi: Ano hito ni wa sootoo sekinin ga atta to omou **naa**.
           '(I) think that guy was pretty much to blame **naa**.'

While insisting on his interpretation of the protagonist's behavior with **sa**, Hiroshi finally gives an assessment adding **na**, which strongly emphasizes the speaker's stance with the flavor of '<u>I</u> think'.
    In sum, each particle serves to signal the function that the utterance preceding the particle has at the particular sequential point in the gossip discourse.

(5) The clusters of particles
    In the following sections, I will point out how Hiroshi and Sachiko use each particle to influence each other's stance toward the protagonist, Takao. There are several 'clusters' of the same particles used by one or both speakers across several utterances in various places in the discourse. At the beginning, the 'authority' particle **yo** and the 'insisting' particle **sa** are actively used by both Hiroshi and Sachiko in contrasting their stance differences toward Takao, Hiroshi being critical,

Sachiko being rather sympathetic, as we see in Example 6. This exchange occurs right after Sachiko has finished telling a story she heard from Takao:

#6
1 Hiroshi:Demoo hi wa yappari okusan to socchi no hoo ni aru daroo **naa**.
　　　　'However, (I) still think that the wife and that side (=Takao) are to blame.'
2 Sachiko:Deshoo?
　　　　'(You think so, too,) right?'
3 　　　:A Takao-san?
　　　　'Oh, (you mean) Takao?'
4 Hiroshi:Takao-san ni aru daroo.
　　　　'Takao is to blame.'
5 Sachiko:((Laugh))
6 Hiroshi:Arunja nai no?
　　　　'Isn't (he)?'
7 Sachiko:((Laugh))
8 　　　Demo **sa**,
　　　　'But **sa**,'
9 　　　okusan ga mainichi onigiri o motte ku kurunda tte.
　　　　'the wife brings riceballs every day, (he told me).'
10 　　　Takao-san wa "iyada" tte ittanda tte **yo**.
　　　　'Takao said "No thanks", (he told me) **yo**.'
11 　　　"Dakedo soredemo motte kurunda."
　　　　'"But (she) still brings them (to me)."'
　　　　(Takao said)
12 　　　Demo yappari Takao-san warui no kashira.
　　　　'But is Takao still to blame, (I) wonder.'
13 Hiroshi:Sore wa waruinja nai yappari.
　　　　'(He) is still to blame, (don't you think).'
14 　　　Aite wa hitozuma dakara **sa**.
　　　　'The other person is someone else's wife **sa**.'
15 Sachiko:Jaa [doosuru]
　　　　'Then how could (he)'
16 Hiroshi:[Sorenari] no atsukaikata attanja nai?
　　　　'There should have been a way to handle (her), (don't you think)?'
17 　　　Kanpeki ni kyozetsu suru toka **sa**.
　　　　'To refuse (her) completely, for example **sa**.'
18 Sachiko:Demo kurasumeeto de mainichi kao awaserunda tte **yo**.
　　　　'But (they) are classmates and meet every day, (he told me) **yo**.'
([ ] indicates overlap)

Hiroshi clarifies his stance with **na** in line 1. Sachiko and Hiroshi's stance differences become clear when Sachiko gives him a 'dispreferred' response of laughter instead of a quick response of agreement in line 5. In lines 8, 9 and 10, Sachiko tries to convince Hiroshi by using the information she heard from Takao. She attracts Hiroshi's attention with utterance-internal **sa**, and with the authority particle **yo**. In

line 13, however, Hiroshi does not change his stance. He further warrants his claim that Takao is to blame by using **sa** in line 14. Hiroshi uses **sa** in suggesting an alternative action Takao could have taken in line 17. In line 18, Sachiko again uses information from Takao in conveying her stance that Takao is a passive victim.

Although they both agree that the wife, Mariko, is responsible for what happened, they have different stances toward her husband's action of putting pressure on Takao through the administration to quit school. Hiroshi thinks it is quite a natural and understandable action for a Japanese husband, whereas Sachiko thinks it is extreme to involve the school administration in a private matter. Sachiko tries to pull Hiroshi's stance closer to hers. What we see in Example 7 is her switch from 'agreement seeking' **ne** to 'insisting' **sa** clusters:

#7

1 Sachiko:Soo **nee**.
        'Well **nee**.'
2   Nan te iu ka,
         'What shall (I) say,'
3   gakkoo kara yamesasero tte itte kuru no wa chotto ikisu- ikisugi tte iu ka **ne**,
         '(I think) it is too radical that (the husband) told (Takao) to quit school via the administration **ne**,'
4   uuuun to,
         'uhm,'
5   gakkoo o tooshite Takao-san no gakugyoo o tsuzukerarenai yoo ni suru tte iu yori mochotto mae ni anoo-
         'before preventing Takao from continuing school through the administration, uh-'
6 Hiroshi:Doko de?
         'At which point?'
7   Doko de shori suru?
         'At which point should (the husband) take care of (the matter)?'
8 Sachiko:Nnnnn.
         'Hm.'
9   Iya dakara **sa**.
         'Well **sa**.'
10   Feisu tuu feisu de **sa**.
         '(he should have dealt with Takao) face-to-face **sa**.'
11   Hanashiatte.
         'By talking.'
12   Hanashiatte tte iu ka,
         'By talking or rather,'
13   donarikomu no wa jiyuu da to iu ka,
         '(I think) (the husband) has a right to come up to (Takao) furiously,'
14   seetoo da to omou, watashi wa.
         'that is justifiable, I think.'
15   Watashi dattara to omou kedo **sa**.
         'I would, I think **sa**.'

In lines 1 and 3, Sachiko uses **ne** to seek agreement, and in line 9, she switches to **sa** to push her point that what the husband did is too radical.

Finally, a few turns after Example 7, Hiroshi wraps up the discussion by asking Sachiko whether she thinks the husband has the right to blame Takao. She says he does. Then Hiroshi confirms Sachiko's stance toward Takao. What we see in Example 8 is the 'agreement' particle **ne** clustering in the wrap-up portion of the gossip discourse:

#8

1 Hiroshi:De soo iu kanari wa aru to omou.
　　　　　'So (you) think so (=Takao is very much to blame).'
2 Sachiko:Un.
　　　　　'Yeah.'
3　　　　 Hi wa aru to omou.
　　　　　'(I) think (Takao) is to blame.'
4　　　　 Soo iu jookyoo o tsukuchattanda kara.
　　　　　'Because (he) has created such a situation.'
5 Hiroshi:Un.
　　　　　'Yeah.'
6　　　　 Soo desu **nee**.
　　　　　'That's right **nee**.'
7　　　　 Kanari hi wa aru **nee**.
　　　　　'(Takao) is pretty much responsible for (what happened), isn't (he) **nee**.'
8 Sachiko:Un.
　　　　　'Yeah.'
9 Hiroshi:Dakara .. gakkoo o tooshite **ne**,
　　　　　'So .. through the school **ne**,'
10　　　　 maa **ne**,
　　　　　'well **ne**,'
11　　　　 soo iu ... are suru no wa ikisugi kamo shinnai kedo **nee**.
　　　　　'to do that (the husband putting pressure on Takao) may be too radical **nee**.'
12 Sachiko:Aaa.
　　　　　'Yeah.'
13 Hiroshi:Danna ni shitara **nee**.
　　　　　'But for the husband **nee**.'
14　　　　 Datte mentsu marutsubure dashi **nee**.
　　　　　'(He) completely lost face (don't you think) **nee**.'
15 Sachiko:Aa, soo **nee**.
　　　　　'Yeah, that's right **nee**.'

In lines 3 and 4, Sachiko finally accepts Hiroshi's view that Takao is responsible for what happened. Now Hiroshi takes the initiative in this wrap-up portion. He starts using the 'agreement' particle **ne** frequently towards Sachiko, a defeated defender of Takao. In lines 6 and 7, Hiroshi confirms what Sachiko has said by showing agreement using **ne**. From lines 9 to 11, Hiroshi even shows a compromise in his view toward the husband's action. He takes into account Sachiko's stance that the

husband's action was too radical. Then in lines 13 and 14, Hiroshi makes comments from the husband's perspective and uses **ne** to seek Sachiko's agreement on his comments. Sachiko agrees with him, using **ne**.

This **ne** cluster shows that Sachiko accepts Hiroshi's stance towards Takao, and Hiroshi accepts Sachiko's stance toward the husband's action. In other words, the cluster of the 'agreement' particle **ne** in the wrap-up portion indicates that the participants' stance differences are finally resolved.

In summary, three sections of particle clusters have been discussed. One is a **yo** and **sa** cluster found where the participants' stance conflict becomes clear, one is a switch from **ne** to **sa** when Sachiko tries to convince Hiroshi of her stance, and finally, we see a **ne** cluster towards the end, where the stance differences are resolved.

(6) The overall flow of the gossip and particle clusters

The chart in the Appendix shows the overall flow of this discourse and the particle clusters found in the data. Immediately after Sachiko finishes telling the story of Takao (A), Hiroshi and Sachiko start exchanging assessments of Takao, Mariko, and her husband, going back and forth. First, they try to influence each other's stances by using **sa** and **yo**, as we have seen in example #6 (B). Then, both of them use **ne** frequently when they talk about the husband's action (C). This cluster shows that Sachiko pushes her stance, seeking agreement from Hiroshi, while Hiroshi uses **ne** to show acceptance, indicating 'I see what you are saying.' Hiroshi and Sachiko not only express their assessments straightforwardly, but also talk about hypothetical situations, putting themselves in these protagonists' shoes (F). Hiroshi uses **sa** and **yo** frequently and criticizes Takao. Then, in order to defend the husband's action, he uses **sa** 12 times (G) both utterance-internally to keep Sachiko's attention, and utterance-finally to push his argument forward. Then Sachiko expresses her critical stance toward the husband using **ne** and then **sa**, as in Example #7 (I). After all these pushing and pulling uses of the particles, their stance conflict is resolved in the final cluster of **ne**, as shown in Example #8 (J).

III. Concluding remarks

These particles, interacting with other linguistic and non-linguistic cues, convey the speakers' attitudes towards their utterances. Further, as the clusters show, the particles serve the function of negotiating stance differences between Hiroshi and Sachiko. As Oishi (1985) and Cook (1988) noticed in their natural interactional data, particles play an important role in constituting discourse in which both participants try hard not only to communicate their own stance, but also challenge their interlocutor's stance.

More broadly, particles in this gossip lead Hiroshi and Sachiko to agreement about the ethics of this situation involving a married couple. Hiroshi and Sachiko, who are members of a small Japanese student community in the States, use the particles strategically to negotiate their views on what is acceptable behavior and what is not for a member of the community.

Brenneis (1984:490) mentions that 'a crucial way of demonstrating one's membership is through sharing what is "common knowledge" in the community -- what "everyone" knows.' When Hiroshi makes a substantial generalization ('Japan

is that kind of society **sa**'), he claims his membership in the community by making his opinion sound like what 'everyone in the community knows.' And by tacitly claiming his membership, he imposes his way of thinking on Sachiko with the particle **sa** and tells her 'That's how you are supposed to think.' In other words, Hiroshi exerts power over Sachiko by evoking his membership in the community.

On another level, the particles serve to maintain the social relationship between Hiroshi and Sachiko. This gossip discourse can be viewed as an example of 'sociable argument' (Schiffrin 1984). Hiroshi and Sachiko are good friends to begin with, and in the gossip, they 'play with' the tension caused by their different views toward Takao and the husband. By insisting on their own stance or contrasting their stances with one another, they put distance between each other. And by seeking or showing agreement, they come close to each other again. In other words, the particles occurring throughout the gossip discourse play an important role in negotiating the participants' social distance as well as their stance differences.

References

Brenneis, Donald. 1984. Grog and gossip in Bhatgaon: style and substance in Fiji Indian conversation. American Ethnologist 3.487-506.

Clancy, Patricia M. 1985. The acquisition of Japanese, ed. by D. I. Slobin, 373-524. Hillsdale: Erlbaum.

Cook, Haruko. M. 1988. Sentential particles in Japanese conversation: a study of indexicality. Ph.D. dissertation. University of Southern California.

Grimshaw, Allen (ed.) in press. Conflict talk. Cambridge: Cambridge University Press.

Kuno, Susumu. 1973. The structure of the Japanese language. Cambridge: MIT Press.

Martin, Samuel E. 1975. A reference grammar of Japanese. New Haven: Yale University Press.

Oishi, Toshio. 1985. A description of Japanese final particles in context. Ph.D.dissertation. University of Michigan.

Schiffrin, Deborah. 1984. Jewish argument as sociability. Language in Society 13, 311-335.

Schiffrin, Deborah. 1985. Everyday argument: the organization of diversity in talk, ed. by Teun A. van Dijk, 35-46. Orlando: Academic Press.

Schiffrin, Deborah. in press. The management of a cooperative self in argument: the role of opinions and stories, ed. by Allen Grimshaw. Cambridge: Cambridge University Press.

Seki, Minako. 1987. Final particles in the speech of 2-year-old Japanese children and their mothers. ms. University of Southern California.

Slobin, Dan. I. (ed.) 1985. The crosslinguistic study of language acquisition. Hillsdale: Erlbaum.

Tsuchihashi, Mika. 1983. The speech act continuum: an investigation of Japanese sentence final particles. Journal of Pragmatics 7. 361-387.

Uyeno, Tazuko Y. 1971. A study of Japanese modality: a performative analysis of sentence particles. Ph.D. dissertation. University of Michigan.

van Dijk, Teun (ed.) 1985. Handbook of discourse analysis, III: discourse and dialogue. Orlando: Academic Press.

Yamada, Akiko. 1980. The development of syntactic devices for communication in the early acquisition of Japanese: a case study. Master's Thesis. University of Hawaii.

Appendix
The stream of gossip and clustering of particles

	Hiroshi	Sachiko
A Sachiko tells Hiroshi the story of Takao		
B Assessment of Takao (Example #6)	sa (11) yo (4)	yo (3)
C Assessment of the husband	ne (6)	ne (5)
D Assessment of Takao		
E Assessment of Mariko		
F Assessment of Takao (What Hiroshi would do)	sa (5) yo (3)	yo (3)
G Assessment of the husband	sa (12)	
H Takao would do the same as the husband in Japan		
I Assessment of the husband (Example #7)		ne (3) sa (3)
J Assessment of Takao (Example #8)	ne (7)	ne (1)

Information Structure and the Scope of Sentential Negation*

Enric Vallduví
University of Pennsylvania

1   Introduction

Apparent irregularities in the semantic scope of sentential negation have been noticed by many linguists and philosophers. A well-known problem is the existential force of most sentential subjects in negative statements. Many authors conclude from the presence of this existential force that the subject lies outside the scope of negation (e.g. Frege 1892, Strawson 1950, Kamp 1981). Another irregularity, noticed by Kraak 1966, Jackendoff 1972, Gabbay & Moravcsik 1978, Kuno 1980, Payne 1985, McGloin 1987, and Horn 1989, among others, is that, in certain contexts, some parts of the predicate do not seem to be affected by sentential negation either. It has been suggested that the scope of negation must exclude these 'non-negated' constituents as well. A different approach to the first problem (e.g. Kempson 1975, Gazdar 1979, and Horn 1986), or to both (Horn 1989), however, maintains that sentential negation is indeed external and has scope over the entire subject-predicate structure, and that the existential force of subjects and the feeling that only part of the predicate is negated in certain sentences are better handled by a non-truth-conditional account. Similar observations are applicable to the yes/no question operator or Q-operator, as discussed in Kuno 1980, 1982.[1]

This paper provides a unified account of these two phenomena. Following Horn's externalist point of view, and using evidence from Catalan and English, it is argued that the readings where part of the sentence is felt as 'non-negated' are the outcome of the interaction of logico-semantic structure and informational structure, the latter being the component of language where 'packaging' relations like focus/open-proposition (Prince 1981, 1986, Ward 1985) are established and interpreted. In particular, it is proposed that informational meaning and semantic meaning interact by means of partial cancellation, yielding the understandings where some part of the clause is felt to escape the scope of negation or the Q-operator, without the need to resort to any additional constraints.

2   The Facts

In what follows, we will refer to the readings where part of the proposition is felt to lie outside the scope of negation or the Q-operator as INFRAPROPOSITIONAL READINGS, and to the elements that lie outside the scope of negation or the Q-operator as OUTSIDER TERMS. As we have seen, outsider terms may be either subjects or complements of V.

2.1 Subjects

The scopal relation between subject definite descriptions and sentential negation has been a matter of controversy for many years. A sentence like (1), in one of its 'readings',

(1)   Gomez Addams didn't sell his South American holdings.

seems to convey the understanding that Gomez Addams exists; in other words, that there is a Gomez Addams about whom we communicate a negative predicate. Some semantic theories incorporate this understanding in their formal semantic representation of negation by either arguing or assuming that negation is internal and does not have scope over the subject (Frege 1892, Strawson 1950, Kamp 1981, inter alia) or by arguing that the subject in (1) must be analyzed as an existential quantifier which gives rise to logical scope interactions with the negation operator (Russell 1905) (cf. Lukasiewicz 1922 for a different ambiguist approach), yielding the wide-scope reading of negation, where there is no existential claim for *Gomez*, and the infrapropositional reading, where there is.

In contrast, one may address the issue in a totally different fashion: let negation have scope over the entire predication, including the subject, at all times, and attribute the existential force of the subject NP in (1) to some non-truth-conditional property of this NP (cf. Kempson 1975, Gazdar 1979, Horn 1989, among others).[2] Horn (1989:§7.3.4), for example, argues that subjects tend to be felt as lying outside the scope of negation because they tend to be topics. Topics, being what the sentence is about (cf. Strawson 1964, Reinhart 1981) are not within the scope of assertion and, therefore, are pragmatically understood as being outside the scope of negation as well.

2.2 Complements of V

It is not subjects alone which appear to give rise to infrapropositional readings. Jackendoff (1972) notices that sometimes negation seems to apply to only part of the predicate (cf. also Gabbay & Moravcsik 1978). One such example is, for instance, the sentence in (2) (in Jackendoff 1972:255),[3]

(2)   Max didn't kill the JUDGE with a hammer.

where negation is felt to affect only the direct object and not the entire VP or sentence, i.e., not only the subject *Max* but also the PP *with a hammer* behave like outsider terms. Jackendoff argues that this is an example of his rule of 'association with focus', a rule that somehow connects logical operators with the focus element in the sentence to establish an intimate tie between the two.

Kuno (1980, 1982) notes similar restrictions for negation or the Q-operator in Japanese: their scope generally only extends to the focal verbal constituent that immediately precedes them. In fact, English, while distinct from Japanese in word order and directionality of scope, also shows parallel effects for both operators. Compare (2) with (3). Notice that in (3) only *judge* is understood as the 'aim' of the yes/no question.

(3)   Did Max kill the JUDGE with a hammer?

In most of the literature, it is clearly concluded or tacitly assumed that the semantics of sentences like (2) and (3) cannot include a sentential negation operator with scope over the entire proposition. When trying to define the logico-semantic scope of negation in these sentences, however, we encounter references to non-logico-semantic, informational notions like focus and theme. A clear example is Payne (1985), who does not seem to clearly endorse an exclusively logico-semantic approach to the diversity of readings, and affirms, using Praguean terminology, that 'the contextually bound elements are removed from the scope of negation, and what is actually negated is the contextually free portion of the sentence' (1985:199). And it is again Horn (1989:515) who proposes that there should be no need to resort to multiple logico-semantic ambiguity to account for the VP infrapropositional readings: 'the negative element takes semantic scope over the entire predication, but [...] will be understood as associated with that rhematic constituent which receives the intonation peak'.[4]

3   Informational Meaning

It is precisely this position, expressed by Payne and Horn with respect to VP outsider terms, that we develop in this paper in terms of the interaction of semantic and informational meaning, while applying it to all outsider terms: subjects and complements of V. We take the position that sentential negation and the Q-operator always have semantic scope over the entire predication, with no exceptions, and we derive the infrapropositional readings from the different informational structures of the sentence and their different overlapping patterns with the invariable semantic structure. Before moving on, however, we must introduce some theoretical notions regarding informational meaning and the informational component.

3.1   What is Informational Meaning?

It is well known that sentences that are truth-conditionally equivalent can be encoded in different syntactic constructions. This logico-semantically-vacuous variation has been attributed to differences in informational meaning. Informational meaning represents a packaging of information in discourse (Chafe 1976, Prince 1986) that reflects the beliefs of the speaker about the hearer's attentional state. Thus, informational meaning is composed of a set of instructions with which speakers direct hearers to enter the information encoded in a given sentence into their discourse model or, alternatively, their stock of knowledge (cf. Vallduví, forthcoming). Informational meaning is established and interpreted in the informational component of language (cf. Lambrecht's 1988 'information structure'), which is linked to the syntax in the same way the logico-semantic component is.

Informationally, we divide the sentence into **focus** and **ground** (S={ **focus, ground**}), and the latter is subdivided into the **link** and the **tail** (**ground**={ **link, tail**}). The focus constitutes the assertion of the utterance in von Stechow's (1981) and Lambrecht's (1988) terms, or the 'informative part' in Halliday's (1967),

and it is the only non-elidable part of the sentence, since it is the only addition of information at the time of utterance. The ground roughly corresponds to the open-proposition or presupposition in Akmajian 1979(1970), Chomsky 1971, Prince 1981, 1986, or Lambrecht 1988. The ground is treated as already relevant and 'around' somehow at the time of utterance. Its main informational force consists in anchoring the focus appropriately in the discourse model or stock of knowledge. It is further divided into the link, which corresponds to (a subset of) the topic in the topic-comment framework (or the theme in Halliday 1967), and the tail, which bears some overlap with what has been called antitopic (Chafe 1976, Lambrecht 1981). The former is always sentence-initial and functions as an address-marker, in the sense that it activates a given discourse entity or a given address in the stock of knowledge under whose label the oncoming information is catalogued (hence the aboutness feeling).[5] Further detail on these informational primitives can be found in Vallduví, forthcoming.

3.2 Structural Representation

Obviously, informational meaning must be recoverable from the surface shape of the sentence. This can be done by morphological, phonological or syntactic means, or, as is usually the case, by means of a combination of these. In Catalan —to a larger extent than in English— the informational articulation of the sentence is structurally expressed in the syntax by means of right- or left-detachment of non-focal constituents, yielding the configuration in (4) (cf. Vallduví 1989):

(4)　　$[_{XP}$ non-focus (link) $[_{IP}$ FOCUS] non-focus]

This abstract configuration is instantiated in sentences like the ones in (5). The example in (5)a is a right-dislocation, or *emarginazione* in the Italian literature, and (5)b is a left-dislocation (functionally equivalent to an English topicalization):

(5)　　a. $[_{XP}$ $[_{IP}$ hi$_i$ VAN OBRIR　BOTIGA,] a Londres$_i$.]
　　　　　　　　　loc PAST-3p-open store　　　in London
　　　　　　'(They) OPENED A STORE in London.'

　　　　b. $[_{XP}$ A Londres$_i$ $[_{IP}$ hi$_i$ VAN OBRIR　BOTIGA.]]
　　　　　　　in London　　　*loc PAST-3p*-open store
　　　　　　'In London (they) OPENED A STORE.'

　　　　c. cf. the canonical:
　　　　　　(*Hi) van obrir botiga a Londres.

In (5), the presence of the clitic locative pronominal *hi* (compare the cliticless canonical in (5)c, string order in (5)b, and prosody in (5)a show that the locative PP *a Londres* is in an external non-argument position.[6] In addition, Catalan being a null-subject language, it has been argued that whenever overt preverbal subjects occur, they are not in a canonical subject slot in the clause, but in an external adjunction position (Rigau 1988; cf. Bonet 1989 for a similar analysis). In other words, the

position of preverbal overt subjects, like *els Lladró* 'the Lladró Brothers' in (6)a or (6)b,

(6) a. [$_{XP}$ Els Lladró$_i$ [$_{IP}$ *pro$_i$* VAN OBRIR   BOTIGA A LONDRES.]]
                       PAST-3p-open store   in London
    'The Lladró Brothers OPENED A STORE IN LONDON.'

   b. [$_{XP}$ Els Lladró$_i$ [$_{XP}$ [$_{IP}$ *pro$_i$* hi$_j$ VAN OBRIR   BOTIGA] a Londres$_j$.]]
                        loc PAST-3p-open store   in London
    'The Lladró Brothers OPENED A STORE in London.'

   c. [$_{XP}$ A Londres$_j$ [$_{XP}$ [$_{IP}$ *pro$_i$* hi$_j$ VAN OBRIR BOTIGA] els Lladró$_i$.]]
    'In London the Lladró Brothers OPENED A STORE.'

is equivalent to the position occupied by *a Londres* 'in London' in (5)b or (6)c. The difference between the two is that, while the pronominal bound by the PP must be overt, the pronominal bound by the subject NP is phonologically null (*pro*, in some theories). I'll assume here, following some of the work cited above, that the actual argument position for subjects in Catalan is clause-final, which is the other position in which they appear at the surface.[7] ¿From this discussion we see that subjects in Catalan tend to be links just as they do in English (Horn 1989; cf. Section 2.1 above). The difference is that in Catalan linkhood is marked structurally at the surface, while in English it is not. The same is the case with non-focal non-link material —the tail— which is right-dislocated in Catalan, but just destressed and left at the end of the clause in English.

3.3   Informational Formalization

Using the informational primitives introduced above in this section (§ 3.1), the informational interpretation of, say, sentence (6)b is the following: 'I instruct you to activate or go to the discourse entity 'the Lladró Brothers' and I inform you (or I assert) that 'opening a store' is a satisfactory instantiation of the gap in the relevant assumed frame *the Lladró Brothers stand-in-some-relation-with London*'. This is formalizable with the informational formula in (7):

(7) $\Lambda np_1, np_1 = $ L. Bros., λpp_2 [Φ [*yes*(np_1 **opened a store** pp_2)]](in-London)

where a) the link operator, Λ, is represented as a quantifier-like element in a quantifier-variable structure, b) Φ stands for a one-place focus or assertion operator that takes the clause as its argument, c) the tail elements are lambda-abstracted from the clause, and d) *yes* merely stands for the affirmation operator, tacit in most languages, which is in complementary distribution with the interrogative Q-operator and the negation operator. It must be emphasized that this formula represents informational meaning and not logico-semantic meaning. The interpretation of the packaging instructions represented in (7) takes place separately from but in parallel with logico-semantic interpretation.

   Representations for other informational readings —the list is not exhaustive— of the same logico-semantic proposition are displayed in (8). Using Lambrecht's (1988)

terminology, (8)a is an example of sentential focus, (8)b an example of predicate focus, and (8)c an example of verbal narrow focus:[8]

(8) a. (...quan) VAN OBRIR BOTIGA A LONDRES ELS LLADRÓ.
         '(...when) the L. Bros. opened a store in London.'
         Φ [*yes*(**L. Bros. opened a store in London**)]

     b. Els Lladró$_i$ *pro$_i$* VAN OBRIR BOTIGA A LONDRES. (=(6)a)
         'The L. Bros. OPENED A STORE IN LONDON.'
         $\Lambda np_1, np_1 =$ L. Bros., [Φ [*yes*(np_1 **opened a store in London**)]]

     c. Els Lladró$_i$ *pro$_i$* n$_j$'hi$_k$ VAN OBRIR, de botiga$_j$, a Londres$_k$.
         'The L. Bros. OPENED a store in London.'
         $\Lambda np_1, np_1 =$ L. Bros., $\lambda np_2 \lambda pp_3$ [Φ [*yes*(np_1 **opened** np_2 pp_3)]] (in-L)
         (store)

It is impossible to fully justify here this informational representation, which is obviously empirically modelled after the syntactic representation of informational meaning in Catalan. Suffice here to say that, as opposed to other analyses, our formalization is more efficient in handling instances of sentential or predicate focus, which are far more common than cases of narrow focus, and fares well in cases of double focus too. The fact that the non-focal elements rather than the focus is abstracted away from the clause may seem odd to the reader familiar with representations of focus currently available in the literature, but our characterization reflects the conceptual argument that what is informative or asserted, the focus, is the actual reason for the clause to exist, and, therefore, at the relevant level, the former must be represented as standing within the latter.

4   Deriving the Infrapropositional Readings

4.1   The Negation and Q- Operators

Let us finally move to the goal of this paper: we are now ready to try to obtain the infrapropositional readings in negative and interrogative sentences. As we have seen, the semantic meaning of sentences (6) and (8) has remained constant, while different informational understandings have been established. In other words, the logico-semantic proposition

(9)   *yes*[open (store, in London, L. Bros.)]

is shared by all the sentences in (6) and (8). However, each of these sentences has a different informational representation, some of which we have already spelled out in the above examples. So far, there has been no conflict in the understandings derived from both types of meaning: the interaction between the two has null effects. This, however, is not always the case. To see this, let us extend the informational representations in (7) and (8) to their negative and interrogative counterparts, i.e., sentences where, instead of having a *yes* operator we have negation ($\sim$) or a Q-operator (Q).

Informational representations for the interrogatives are in (10) (*que* is the Catalan instantiation of the Q-morpheme):[9]

(10) a. Que VAN OBRIR BOTIGA A LONDRES ELS LLADRÓ (, pas)?
'Did the L. Bros. open a store in London?'
Φ [Q(**L. Bros. opened a store in London**)]

b. Els Lladró$_i$ que *pro$_i$* VAN OBRIR BOTIGA A LONDRES?
'The L. Bros. did they OPEN A STORE IN LONDON?'
$\Lambda np_1, np_1 =$ L. Bros., [Φ [Q(*np$_1$* **opened a store in London**)]]

c. Els Lladró$_i$ que *pro$_i$* hi$_j$ VAN OBRIR BOTIGA, a Londres$_j$?
'The L. Bros. did they OPEN A STORE in London?'
$\Lambda np_1, np_1 =$ L. Bros., λpp_2 [Φ [Q(*np$_1$* **opened a store** *pp$_2$*)]] (in-L)

d. Els Lladró$_i$ que *pro$_i$* n$_j$'hi$_k$ VAN OBRIR, de botiga$_j$, a Londres$_k$?
'The L. Bros. did they OPEN a store in London?'
$\Lambda np_1, np_1 =$ L. Bros., $\lambda np_2 \lambda pp_3$ [Φ [Q(*np$_1$* **opened** *np$_2 pp_3$*)]] (in-L)
(store)

Sentences (10)a, (10)b, (10)c, and (10)d are informationally equivalent to sentences (8)a, (8)b, (6)b (=(7)), and (8)c, respectively. They have the same informational structure and meaning (aside from the propositional operator[10]), which in Catalan is straightforwardly represented in the surface structure of the sentence. Each of the sentences in (10) shows the same dislocation pattern as its affirmative counterpart in (6) and (8). At the same time, the sentences in (10) are all instantiations of the same logico-semantic proposition in (11):

(11) Q[open (store, in London, L. Bros.)]

We see, then, that while the sentences in (10) are all logico-semantically equivalent, they are, in contrast, informationally distinct from each other.

The same observations can be directly carried over to the negative sentences in (12):

(12) a. (...que) no VAN OBRIR BOTIGA A LONDRES ELS LLADRÓ.
'(...that) the L. Bros. didn't open a store in London.' (neg-external reading)
Φ [$\sim$(**L. Bros. opened a store in London**)]

b. Els Lladró$_i$ no *pro$_i$* VAN OBRIR BOTIGA A LONDRES.
'The L. Bros. didn't open a store in London.' (neg-internal reading)
$\Lambda np_1, np_1 =$ L. Bros., [Φ [$\sim$(*np$_1$* **opened a store in London**)]]

c. Els Lladró$_i$ no *pro$_i$* hi$_j$ VAN OBRIR BOTIGA, a Londres$_j$.
'The L. Bros. didn't OPEN A STORE in London.'
$\Lambda np_1, np_1 =$ L. Bros., λpp_2 [Φ [$\sim$(*np$_1$* **opened a store** *pp$_2$*)]] (in-L)

d. Els Lladró$_i$ no *pro$_i$* n$_j$'hi$_k$ VAN OBRIR, de botiga$_j$, a Londres$_k$.
'The L. Bros. didn't OPEN a store in London.'
$\Lambda np_1, np_1 =$ L. Bros., $\lambda np_2 \lambda pp_3$ [Φ [$\sim$(*np$_1$* **opened** *np$_2 pp_3$*)]] (in-L)
(store)

The four sentences in (12) are informationally distinct from each other. Each of them is informationally equivalent (except for the propositional operator) to the corresponding lettered sentences in (10). This is shown, again, by the Catalan

surface structure. The logico-semantic structure of the sentences in (12), however, remains constant:

(13)   ∼[open (store, in London, L. Bros.)]

This is precisely what Kempson 1975, Gazdar 1979, Horn 1989, and the other partisans of the externalist approach to sentential negation argue for. We will now see how we can derive the infrapropositional readings without fiddling with the logical scope of negation, just as these authors propose. The readings in question are derived by exploiting the partial-overlap interaction between the logico-semantic and the informational representations of sentences like (10) and (12).

Mark that the informational representations for the sentences in (10) and (12) are, obviously, motivated exclusively by the informational meaning they convey, just as they were in the affirmative sentences in (7) and (8). Notice, incidentally, that it is precisely the outsider terms in this sentences which are dislocated to non-argument slots. Catalan clearly reflects structurally the fact that outsider terms must be non-focal, as observed by some of the authors mentioned above.

4.2   A Unified Account

Let us take, for example, sentence (12)b, where the outsider term is the subject. The informational meaning this formula expresses is as follows: I instruct you to activate or go to the discourse entity 'the Lladró Brothers' and I inform you (or I assert) that 'not opening a store in London' is a satisfactory instantiation of the gap in the relevant assumed frame 'els Lladró are-predicated-something-of'. The ground is assumed to be relevant and non-controversial at the time of utterance. While (12)b semantically expresses that it is not the case that the Lladró Brothers opened a store in London, simultaneously it informationally expresses that 'the Lladró Brothers are predicated-something-of' is taken for granted. The subject *els Lladró*, while within the scope of negation in the semantics, remains outside the scope of assertion in the informational representation, and, therefore, as Horn (1989:512) suggests, in some sense, outside the scope of negation as well. This is how we capture his observation about the topichood —linkhood, in our terms— of most subjects. In fact, Catalan postverbal subjects, crucially, are never understood as outsider terms.

Consider now (12)d, for instance: the ground can be described as being 'the Lladró Brothers stand in some relation to a store and to London'. The only informative part, the only actual addition of information to the model at the time of utterance is 'not opening'. Again, while we semantically understand that it is not true that the Lladró Brothers opened a store in London, we informationally understand that some relation holds between them and a store and London. If we cancel out both understandings we obtain that the only element left to be negated de facto is *open*. The informational status of complement-of-V outsider terms, then, just like with their subject counterparts, causes them to be understood as non-negated. This is represented in our informational formula, where only negation and the verb are within the scope of focus.

The same applies to interrogative examples like (10). In (10)c, for instance, in the semantics, *Q* has scope over the entire proposition. But in the informational

component, we interpret that the fact that the Lladró Brothers stand in some relation with London constitutes the ground. The only non-ground material is then 'opening a store', which is left as the only questionable part of the utterance, and, therefore, understood as the aim of the yes/no question.

Notice that with our analysis we directly incorporate Jackendoff's 'association with focus' with no explicit stipulation of it. In fact, Jackendovian examples like the one in (2) are accounted for by our formalism, as shown in (14), or its corresponding interrogative (15):

(14) Max didn't kill the JUDGE with a hammer.
$\Lambda np_1, np_1 = M, \lambda v_2 \lambda pp_3$ [Φ [$\sim(np_1 v_2$ **the judge** $pp_3)$]] (with-hammer) (kill)

(15) Did Max kill the JUDGE with a hammer.
$\Lambda np_1, np_1 = M, \lambda v_2 \lambda pp_3$ [Φ [$Q(np_1 v_2$ **the judge** $pp_3)$]] (with-hammer) (kill)

Again, the propositional operator takes wide logico-semantic scope in both examples, yielding the logico-semantic understandings that it is not the case that Max killed the judge with a hammer for (14) and that the speaker inquires whether or not it is the case that Max killed the judge with a hammer for (15). Both sentences, however, are informationally equivalent: we are informed that *the judge* is an appropriate complement of the ground, where the ground consists of Max standing in a relevant relation with the act of killing with a hammer. Given this, we are taken to understand *the judge* as the 'object' of negation or the aim of the yes/no question.

Crucially, the same treatment is given to subjects with existential force and to VP-internal outsider terms. There is no need to resort to two different sorts of explanation. The trick here is in placing the sentential operator (affirmation, negation, or Q) inside the scope of Φ. This move has null effects when the operator is affirmation, as desired, but important ones when the operator is negation or Q. In particular, it provides us with the infrapropositional readings, i.e. Jackendoff's association with focus, with no additional stipulation of such a rule.

4.3 A Problem

While this approach has been successful in giving a unified account of the existence of outsider terms in negative and Q-operator sentences, it is unable to deal, at this moment, with sentences where sentential negation is clearly a part of the ground, i.e. where it is not 'associated with focus' in any of the ways discussed above. Such is the case in sentences like (16), where (16)b is a literal rendering into Catalan of (16)a (which, apparently, is dialectally restricted in English):

(16) a. My CAR, I haven't paid for yet.
      b. El COTXE, no he pagat encara.
      c. It's my CAR I haven't paid for yet.

In these examples 'not having paid for yet' is the relevant assumed ground which serves to anchor the focus 'car'. Since there is no 'association with focus' it would be

erroneous to try to give an affirmative ground to this sentence. In these sentences, then, we wish to have negation outside the scope of Φ. The very trick that allows us to derive the infrapropositional readings, however, excludes these 'negation-in-ground' readings. Therefore, we need to abstract negation away from the clause, in the same way we would lambda-abstract other tail elements in (16) like *pay*. Before such a step is taken, however, we must determine what repercussions it may have for the overall proposal, and we must investigate the possibility of its being applicable to similar phenomena in the representation of informational meaning.

5 Conclusion

The semantic analysis of sentential negation as an operator with scope over the entire proposition has encountered two major difficulties: the existential force of most subjects and the outsider-term nature of some predicate-internal phrases. It has been concluded from this that in sentences with these irregularities we must abandon the idea of a wide-scope sentential operator.

I have argued that, to the contrary, these long-observed irregularities in the semantic scope of sentential negation need not be incorporated into the semantic representation of negative sentences, but rather, that they follow independently from the interaction of parallel but distinct simple semantic and informational representations. I have provided an independently-motivated formal representation of informational meaning which, when applied to interrogative and negative sentences, duly captures the infrapropositional readings in question with no need of additional rules like 'association with focus'. Finally, and more importantly, the inclusion of non-logico-semantic notions like focus in the semantic representation is rendered unnecessary.[11]

NOTES

* Several conversations with Megan Moser helped this paper take shape. The remaining deformities are my sole responsibility.

1. But see Horn (1989:472-73) for arguments that the Q-operator and negation are not analogous.

2. There are two analyses of this wide-scope negation operator: an Aristotelian two-place subject-predicate operator, or a Fregean one-place propositional operator. This controversy, though important, will be overlooked in this paper. For our purposes here, we will adopt a Fregean propositional operator (but cf. Horn 1989 for compelling arguments for the Aristotelian two-place term operator, which he heralds as superior to its alternatives).

3. In this example, and throughout the paper, capitalization signals the focus of the utterance.

4. The disadvantages and shortcomings of the internalist and ambiguist approaches to infrapropositional readings are discussed in detail in Horn (1989).

5. Herring (1990, this volume), in a comprehensive crosslinguistic study of the syntactic position of informational elements, concludes that contrastive topics —our links— are invariably found in sentence-initial position. This finding is in perfect agreement with our model.

6. Notice, incidentally, that the only way to tell the ungrammatical 'cliticful' string in (5)c from the grammatical string in (5)a is by means of prosody.

7. In spite of this, for the sake of clarity and tradition, and since it is not crucial here, I will still place *pro* in preverbal position in the examples throughout the paper.

8. (8)a is presented as an embedded sentence to improve its felicity. In matrix sentences where two or more complements of V are overt and non-clitic, postverbal subjects are quite marginal. Also, notice in (8)c that the dislocated direct object is binding a clitic back in the clause, just like the locative, which we have already mentioned; in this case, it is a 'partitive' *n*-clitic, with *n, en* and *ne* as allomorphs.

9. I am aware that the English translations in (10) are sort of marginal. I decided to sacrifice English naturalness to obtain a more faithful rendering of the Catalan original. The English preposed phrase should not be read as a 'hanging topic' or full-fledged English left-dislocation (LD-2 in Prince's terms), but as an actual topicalization in a non-interrogative sentence.

10. It is unquestionable that the propositional operator is needed in the informational representation. Not only do we need to derive the infrapropositional readings we are discussing in this paper, but we also have to informationally represent sentences where the affirmation/negation operator is the only focal element:

    (i) She DID pass.
    (ii) She had to pass, and pass she DID.

    In these sentences the operator *yes* is the focus, and it must be represented as such in an informational formula by positioning it within the scope of Φ. These sentences are discussed in detail in Ward 1985 and also in Prince 1986.

11. This type of 'non-mingling' interaction between the different components of language that the proposal presented in this paper argues for is in agreement with the position found in Autolexical Theory and in some recent interpretations of Government-Binding Theory.

REFERENCES

Akmajian, A. 1979 (1970). *Aspects of the Grammar of Focus in English.* New York: Garland.

Bonet, E. 1989. Postverbal Subjects in Catalan. MIT. Unpublished ms.

Chafe, W.L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In C. Li, ed. *Subject and Topic*. NY: AP. Pp. 25-55.

Chomsky, N. 1971. Deep Structure, Surface Structure, and Semantic Interpretation. In D. Steinberg & L. Jacobovits, eds. *Semantics*. Cambridge: Cambridge University Press. Pp. 183-216.

Frege, F. 1892. On Sense and Reference. Translated and reprinted in P.T. Geach & M. Black, eds. 1952. *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell. Pp. 56-78.

Gabbay, D. & J. Moravcsik. 1978. Negation and Denial. In F. Guenther & C. Rohrer, eds. *Studies in Formal Semantics*. Amsterdam: North-Holland. Pp. 251-65.

Gazdar, G. 1979. *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.

Halliday, M.A.K. 1967. Notes on Transitivity and Theme in English (Part II). *Journal of Linguistics* 3.199-244.

Herring, S. 1990 (this volume). Information Focus as a Consequence of Word Order Type. *BLS* 16.

Horn, L. 1986. Presupposition, Theme and Variations. *Parasession Papers. CLS* 22.168-192.

Horn, L. 1989. *A Natural History of Negation*. Chicago: University of Chicago Press.

Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Kamp, H. 1981. A Theory of Truth and Semantic Representation. In J. Groenendijk et al., eds. *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre Tracts. Pp. 277-322.

Kempson, R. 1975. *Presupposition and the Delimitation of Semantics*. Cambridge: Cambridge University Press.

Kraak, A. 1966. *Negatieve Zinnen*. Hilversum: W. de Haan.

Kuno, S. 1980. The Scope of the Question and Negation in Some Verb-Final Languages. *CLS* 16.155-69.

Kuno, S. 1982. The Focus of the Question and the Focus of the Answer. *Parasession Papers. CLS* 18.134-157.

Lambrecht, K. 1981. *Topic, Antitopic and Verb Agreement in Non-Standard French*. Amsterdam/ Philadelphia: John Benjamins.

Lambrecht, K. 1988. When Subjects Behave Like Objects: A Markedness Analysis of Sentence Focus Constructions Across Languages. Expanded version of a paper read at the 1987 LSA Meeting, San Francisco.

Lukasiewicz, J. 1922. On Determinism. Translated and reprinted in S. McCall, ed. 1967. *Polish Logic 1920-1939*. Oxford: Clarendon. Pp. 19-39.

McGloin, N.H. 1987. The Role of *Wa* in Negation. In J. Hinds et al., eds. *Perspectives on Topicalization: The Case of Japanese 'Wa'*. Amsterdam/Philadelphia: John Benjamins. Pp. 165-184.

Payne, J. 1985. Negation. In T. Shopen, ed. *Language Typology and Syntactic Description I: Clause Structure*. Cambridge: Cambridge University Press. Pp. 197-242.

Prince, E.F. 1981. Topicalization, Focus-Movement, and Yiddish-Movement: A Pragmatic Differentiation. *BLS* 7.249-64.

Prince, E.F. 1986. On the Syntactic Marking of Presupposed Open Propositions. *Parasession Papers. CLS* 22.208-22.

Reinhart, T. 1981. Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27.53-94.

Rigau, G. 1988. Strong Pronouns. *Linguistic Inquiry* 19.503-511.

Russell, B. 1905. On Denoting. *Mind* 14.479-93.

von Stechow, A. 1981. Topic, Focus and Local Relevance. In W. Klein & W. Levelt, eds. *Crossing the Boundaries in Linguistics*. Dordrecht: Reidel. Pp. 95-130.

Strawson, P.F. 1950. On Referring. *Mind* 59.320-44.

Strawson, P.F. 1964. Identifying Reference and Truth-Values. Reprinted in D. Steinberg & L. Jacobovits, eds. *Semantics*. Cambridge: Cambridge University Press. Pp. 86-99.

Vallduví, E. 1989. Focus Constructions in Catalan. In C. Laeufer & T. Morgan, eds. To appear. *Proceedings of the Linguistic Symposium on Romance Languages XIX (1989), OSU*. Amsterdam/Philadelphia: John Benjamins.

Vallduví, E., forthcoming. *The Informational Component of Language*. University of Pennsylvania Ph.D. dissertation.

Ward, G. 1985. *The Semantics and Pragmatics of Preposing*. University of Pennsylvania Ph.D. dissertation.

Misbehaving Reflexives in Korean and Mandarin*

David A. Zubin, Soon Ae Chun and Naicong Li
State University of New York at Buffalo

The vast literature on reflexive morphemes and constructions tends to reflect a set of assumptions concerning reflexive phenomena.1 1) There is a tendency to divide uses of reflexive morphemes into distinct types, for example: locally bound, long distance bound, logophoric, point-of-view, intensifying, contrastive, volitional, medio passive, aspectual, etc. 2) Some of these are taken to be conceptual, others to be syntactic, with a sharp boundary between them. 3) The locally bound instances are taken to be central well behaved cases and the more conceptually laden uses are seen as peripheral misbehaving ones. 4) Conceptual uses are claimed to be derivative. For example, logophoric uses are seen as a conceptual analog of the syntactic relation between anaphor and antecedent within the clause.

This paper will argue for a position that in essence reverses the background assumptions given above by providing evidence for the following claims. 1) All uses of reflexive forms in a given language can be shown to be conceptually related within a cognitive semantic framework. Underlying the uses of reflexives in Mandarin and Korean is an abstract indexical system in some ways analogous to the 1st-2nd-3rd person system of interactional discourse. 2) Contrastive and perspectival uses of reflexives most directly reflect this conceptual system and therefore are the most central uses. 3) In narrative discourse there is a continuum from totally free through long distance to clause bound reflexives. The maximally acceptable binding distance reflects degree of identification with the character's psychological perspective. 4) Locally bound reflexives can be viewed as a conventionalized, desemanticized version of perspective.

1. Indexical Systems. The starting point for an analysis capable of integrating the full range of reflexive morpheme uses in Mandarin and Korean is, we argue, the indexical system of Figure 2, which is contrasted for didactic reasons with the familiar indexical system based on the speech situation in Figure 1. Just as 1st, 2nd and 3rd person set up a conceptual polarity indexing the speech situation of interactional discourse, as in Figure 1, Mandarin and Korean have forms which, we propose, constitute an analogous indexical system not based on the speech situation, as in Figure 2. This perspectival system is constituted by an abstract polarity between a center and a periphery, the poles being occupied by abstract actors whom we will call SELF and OTHER.

Forms	speech situation participants
I	speaker
You	intended addressee
S/He	Neither
structure of deictic field	

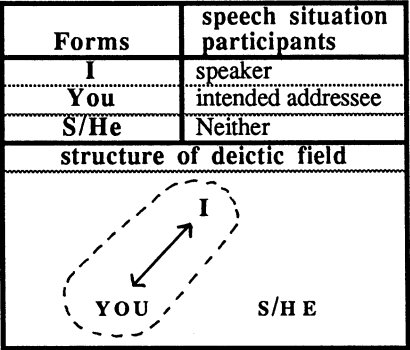| Forms		participants
Mandarin	Korean	
zìjì	caki	SELF
bieren	nam	OTHER(S)
structure of deictic field		

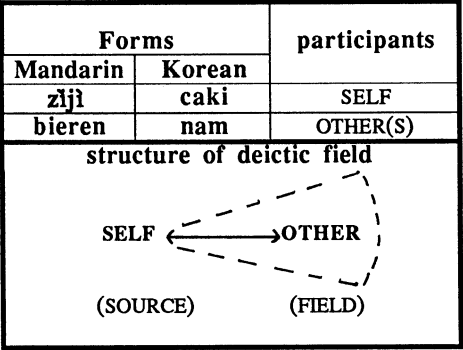Figure 1:
Speech Situation Indexical System

Figure 2:
Perspective Indexical System

The nature of the perspective indexical system can be brought out by specifying some of its properties. First, and crucial for the analysis, it is typically used to set up a contrast between an abstract SELF and OTHER as in the Korean and Mandarin proverbs in (1) and (2)2.

(1) NAM-*uy*   *il-ey sinkyeng-ul nemu ssu-m* CAKI *il-ey*   *soholhakey twointa.*
     OTHER-GEN work-LOC concern-ACC too have-if, SELF work-LOC neglect
     'If one is too concerned about OTHER's business, one would neglect SELF's work.'

(2) *Biérén   de kùnnán jìushì ziji   de kùnnán.*
     OTHER GEN difficulty is   SELF GEN difficulty
     'OTHER's problem is SELF's problem.' (One should care about OTHERs.)

This contrast is inherently perspectival. SELF is at the center or source of the imposed perspective, while OTHER is in the field, or periphery. Note that both SELF and OTHER are referentially non-specific. However, by the Maxim of Relevance SELF and OTHER can be taken to refer to speaker, addressee, or a 3rd person in appropriate context, showing that the perspectival indexical system is not semantically anchored to the speech situation, as is the familiar one of Figure 1. Figure 3 gives a further set of properties which apply assymetrically to SELF and OTHER.

Property	implicational relation
a. referentiality (cf ex 1 and 2)	SELF > OTHER
b.definiteness	SELF > OTHER
c.number (singularity)	SELF > OTHER
d.speech situation participants (I/You) (cf ex 3)	SELF > OTHER
e.focus (cf ex 4)	SELF > OTHER
f. presupposed in implicit contrast (cf ex 4)	OTHER > SELF
g. spatial deictic orientation: at deictic center (cf ex 5 and 6)	SELF > OTHER

Figure 3: Properties of the Perspective Indexical system.

a) In a proverb context such as examples (1) and (2), both SELF and OTHER may be taken to be non-referential, or again by the Maxim of Relevance both may be taken to refer to specific people. But if only one of them is referential, it can only be SELF. This assymetry is given in the implicational relation on the right of the figure. b&c) SELF is more likely to be definite and singular. d) Example (3) shows that SELF is more easily attached pragmatically to 1st or 2nd person of the speech situation. The speaker can easily use (3a) to admonish herself or the listener, but (3b) is incoherent for this purpose.

(3) a. ZIJI   (=> speaker/addressee) *yīnggāi guǎn   hǎo* ZIJI   *de shì.*
    SELF   should manage well SELF GEN business
    'SELF should mind SELF's business.' (=>I/you should mind my/your own business.)

    b. ?? *Biérén* (=> speaker/addressee) *yīnggāi guǎn   hǎo biérén de   shì.*
       OTHER   should manage well OTHER GEN business
    ?? 'OTHER should mind OTHER's business.' (=> I/You should mind my/your own business)

e&f) Example (4) shows that SELF is more easily focused, and OTHER more easily presupposed, than the other way around. In (4a) SELF is focused as subject of the clause, while the contrast with OTHER is presupposed by the focus on SELF. The attempt to reverse these roles in (4b) is incoherent, unless other is taken to be referential ('they should mind their own business' is OK).

(4)a. *CAKI-uy   il-un   CAKI-ka haeya hanta.*
     SELF-GEN work-TOP SELF-NOM do   should
     'SELF should mind SELFs business.' (=> not generic OTHER's business.)

  b. ? NAM-*uy*   *il-un*   NAM-*i*   *haeya hanta.*
     OTHER-GEN work-TOP OTHER-NOM do   should
     ? 'OTHER should mind OTHER's business.' (generic)
        but: They should mind their own business   (specific)

g) Finally, SELF is more likely to be at the spatial deictic center, as shown by coocurrence with the deictic verbs 'come' and 'go' in 5 and 6: (5a) and (6a) are coherent because SELF *goes out from,* and OTHER *comes in to* the deictic center. The (b) versions, which place OTHER at the deictic center, are much less coherent.[3]

(5)a. *CAKI-ka   cikcep chaca-kase* NAM-*eykey chungkohal pilyo-ka epta.*
     SELF-NOM in-person find-go OTHER-DAT advice   need-NOM not be
     'It is not necessary that SELF should go in person to give advice to OTHERS.'

  b. ?*CAKI-ka   cikcep chaca-wase* NAM-*eykey chungkohal pilyo-ka   epta.*
     SELF-NOM in-person find-come OTHER-DAT advice-ACC need-NOM not be
     'It is not necessary that SELF should come in person to give advice to OTHERS.'

(6)a. NAM-tul-i   chaca-wase *CAKI*-lul panghaehal kwenli-ka epta.
     OTHER-NOM visit-come SELF-ACC bother right-NOM don't have
     'OTHERS have no right to come and bother SELF.'

  b. ? NAM-tul-i chaca-gase *CAKI*-lul panghaehal kwenli-ka epta.
     OTHER-NOM visit-go SELF-ACC bother   right-NOM don't have
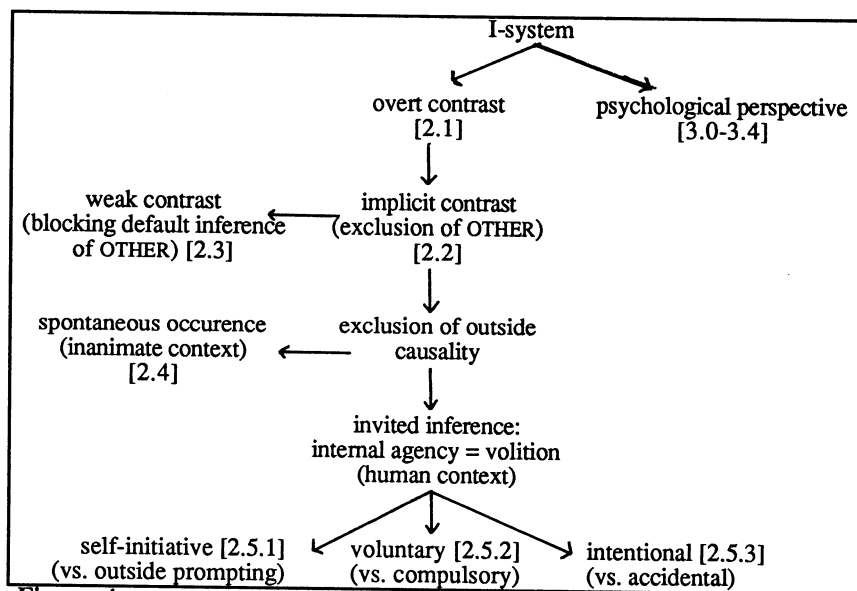   'OTHERS have no right to go and bother SELF.'

Figure 4:
Radial extension of Pragmatic Inferences from the Perspective Indexical System

2.0 Contrast Inferences. Whereas there appear to be a number of independent "homophonic" uses of SELF-forms in Mandarin and Korean, we will argue that they are all inferentially related, and that they form a radial category (Lakoff, 1987) with the perspective indexical system, clearly manifested in the contrastive proverb uses of examples (1) and (2), at its center. The radial relations of the inferences are depicted in Figure 4.4

2.1. Overt Contrast. Examples (1) and (2) above illustrate the use of the perspective indexical system to express an overt contrast between two participants. In these "proverb" contrasts, all aspects of the underlying system are directly manifested: the non-referentiality of the participants, the polarity between them, and the centering on SELF, with OTHER at the periphery.

2.2. Implicit Contrast. The role of the indexical system is further evident when only the SELF pole is actually expressed (the square brackets in (7) and (8) indicate pragmatic inferences). This implicit contrast is a direct inference from the use of the system for overt contrast, as shown in Figure 4. When SELF is mentioned, but not OTHER, OTHER is inferred to be excluded from the event in question. See also (3a) and (4a).

(7)   *CAKI-ka   hal   il-un   CAKI-ka haca.*
      SELF-NOM doing duty-TOP SELF-NOM do
      'SELF should do the duty that belongs to SELF.' [=> not let OTHERS do it].

(8)   *ZIJI   dòng shŏu, fēngyīzúshí.*
      SELF move hand abundant-clothing-abundant-food
      'Rely on SELF's work to get a good life.' [=> not on OTHERS'].

2.3. Weak Contrast. Furthermore, the assertion may not seem contrastive at all when the contrast mechanism is used to defeat a default inference. We illustrate this with actions directed at body parts. In (9), mouth gagging is normally directed by an agent at another person's mouth, as in (9a). This comes through as a default inference in the version of (9a) with a zero posessor. (9b) illustrates the use of SELF to defeat this default inference, since SELF implies the exclusion of an OTHER in the event. Yet there is no intuitive contrast in (9b) beyond the fact that this type of event is normally directed at someone else. (10) provides a parallel example in Korean.

(9)a *Píngping dǔ zhù { tā de / Ø }   zuǐ.*
  Pingping gag stop { s/he GEN / Ø } mouth
  'Pingping$_i$ gagged her$_j$ mouth.'

  b *Píngping dǔ zhù ZIJI   de   zuǐ.*
  Pingping gag stop SELF GEN mouth
  'Pingping$_i$ gagged her$_i$ mouth.'

(10) *Insu-nun { ku / Ø / CAKI}-uy   ttakwi-ul ttaeliessta.*
  Insu-TOP { s/he / Ø / SELF }-GEN cheek-ACC   slap
  'Insu$_i$ slapped { his$_j$ /his$_j$ / his$_i$ } cheek.'

Note that for actions which are normally directed at one's own body part, as in (11), the use of an overt SELF-form gives rise to an inference of conscious control (cf section 2.5.3), since there is no default inference of OTHER to block.

(11) *Píngping zhǎ le yīxià { Ø / ZIJI de } yǎnjīng.*
  Pingping blink PFV once { Ø / SELF GEN} eye
  'Pingping blinked her eyes / Pingping blinked her eyes (consciously).'

2.4. Spontaneous Occurrence5. The next step in Figure 4 is the weakening of a specific outside agent to the vague sense of outside causality. Example (12) shows that a reflexive form may be used with an inanimate actor to exclude generalized outside causality, resulting in an inference of spontaneous occurrence. The inference in (12) is that it wasn't the weather, wind, rain, etc., but that the leaves just drifted down by themselves.

(12) *gānkū de yèzi ZIJI suōsuōde wǎng   xià diào.*
  dried NL leaves SELF rustling towards down fall
  'the dry leaves SELF fell down rustling.'   Lin Jinlan, Zhu. In Li, p.105

2.5. Volitionality. A further invited inference in Figure 4 may be made from the exclusion of outside causality in the context of a human actor, namely, that agency is unexpectedly internal to the actor. This volitional inference has a variety of specific interpretations.
2.5.1. Unprompted Action For example, when visiting someone else's house one normally waits for refreshments, etc. to be offered, rather than just taking them. The default expectation is thus externally prompted action. But the SELF-form can be used to defeat this expectation, i.e. to shift this specific aspect of causality from external to internal. In (13) the SELF-form leads to the inference that the girl Huang Pingping helped herself to the candy. If the host had offered the candy, the SELF-form would be incoherent.

(13) [Huang Pingping has come to visit an old but powerful man whom she knows has a liking for her]

a. *Huáng Píngping shuō   zhe, ZIJI **dǎkāi tánghé**, tiāojiǎn zhe,*
Huang Pingping speaking DUR SELF open candy-box, choose DUR

b. *"Shàngcì lái,   hái yǒu jiǔxīn qiǎokelì ne, zhècì   zěnme méi le?"*
last-time come, still have wine chocolate SFP, this-time why no-more CRS

'(a) While speaking, **Huang Pingping SELF opened the candy box,** choosing among the candies. (b) "Last time when I was here, there were still wine chocolates! How come there isn't any more?"'   *Ye yu Zhou*: p. 587

Example (14) gives another type of social situation. The default expectation is that the older, higher status person will initiate the handshake. If the student Insu initiates the handshake, as in (14a), the SELF-form helps to defeat the expectation of external prompting, i.e. from Professor Kim. But if as in (14b) the handshake is initiated by Professor Kim, the expected initiator, a SELF-form is incoherent.

(14) *Kim-kyosu-wa   Insu-nun keli-eyse macuchiessta.*
Kim-professor-with Insu-TOP street-on met.

a. *Insu-nun   CAKI-ka ap-ulo takaka aksu-lul haessta.*
Insu-TOP SELF-NOM front-to go, handshake-ACC did.
'Insu SELF walked forward and shook hands.'

b. *Kim-kyosu-nun (??CAKI-ka) apulo takaka aksu-lul haessta.*
Kim-prof-TOP   SELF-NOM front-to go, handshake-ACC did.
'Professor Kim (??SELF) walked forward and shook hands.'

2.5.2. Voluntary Action. Related to this is the inference that an activity is voluntary against a background default inference of external compulsion. In (15) the introductory context indicates external force. In the subsequent clause SELF invites the inference that the activity was voluntary. Since the two clauses are connected by a concessive conjunction, only (15b) is an acceptable sequel.

(15) *Usa-ka   Insu-eyekey yak-ul   meki-lyeko haessuna,*   'the doctor tried to
doctor-NOM Insu-DAT medicine-ACC feed-tried-but   feed the medicine to
                                                                                     Insu, but...'

a. ? *Insu-nun ku yak-ul   mekessta.*
Insu-TOP the medicine-ACC ate.   'Insu took it.'

b. *Insu-nun ku yak-ul   CAKI-ka mekessta.*
Insu-TOP the medicine-ACC SELF-NOM ate.   'Insu took it voluntarily.'

2.5.3. Intentional Action. A closely related use of SELF is to defeat the default inference of accidental occurrence, as in (16a). Burning one's fingers normally happens by accident. (16b) shows that adding SELF to the sentence invites the inference of internal causality, in this case intentional action, even though there was no external causality to block. Example (11) provides a parallel case in Mandarin. Figure 5 sums up the volitional inferences.

(16) a. *Youngi-ka   0 sonkalak-ul pul-ey teyessta.*
       Youngi-NOM   fingers-ACC fire-at burned
       'Youngi burned her fingers in the fire (by accident).'

   b. *Youngi-ka   CAKI-ka sonkalak-ul pul-ey teyessta.*
      Youngi-NOM   SELF-NOM fingers-ACC fire-at burned
      'Youngi burned her fingers in the fire (on purpose).'

default inference	resulting inference	example
activity is sociallly prompted	unprompted	13, 14
activity is forced	voluntary	15
activity is accidental	intentional	11, 16

Figure 5: Volitional Inferences evoked by SELF.

3.0. Perspectival Inferences. The second set of inferences derived from the perspective indexical system of Figure 2 revolves around the concept of perspective, especially in narrative discourse. Perspective involves a structuring of the text which invites the listener to identify to some degree with one of the characters of the story world, hereafter called the source character. This identification may involve placing oneself in the spatio-temporal location of the source, or knowing about the mental states of the source, or participating in the source's ongoing perceptual or cognitive processes. One way of characterizing these variations in perspective is the extent to which the listener has access to the mental world of the character. In spatio-temporal perspective the most the listener can do is to infer what the source has perceptual access to, and therefore might be thinking or feeling, hence a low degree of access. In a psychological report, to use Wiebe's (1989) term, the listener is given an external characterization of the source character's perceptions, ideas, opinions, emotions, etc., and through this has partial access to the source's mental world. In represented perception and thought, to use Banfield's (1982) term, the text gives the listener a direct rendition of the Here-and-Now of the source character's actual perceptual and cognitive processes, and therefore full access to the source's mind. A simile may help. In spatio-temporal perspective the presentation of the story world is focused at a spatio-temporal center, but the window into the mind of the source character is nearly opaque. The listener can only infer what is going on inside from the source's behavior. In psychological reports this window into the source character's mind is translucent, and in represented perception and thought the mind window is transparent. As the listener has greater and greater access to the source's mind (as the window becomes more and more transparent), the source character as an objective actor in the story world fades more and more into the background. A textual consequence of this backgrounding is the use of constructions which allow deletion of the source, as in English: "A light was visible in the distance."6

Figure 6 displays the parallel structuring of the indexical system and the concept of perspective. Whereas the contrast-based uses of SELF (sect. 2) exploit the opposition between SELF and OTHER in the system, the perspectival uses disregard OTHER and exploit the epistemic centering on SELF. Note, for example, that when a proverb based on the I-system (such as (1) and (2)) is applied to a specific person, it is SELF, rather than OTHER, that applies to the person, and the whole is taken to have directive illocutionary force, i.e. to involve the intentions, and thus the mental states of the person. Thus SELF is an invitation not only to focus on the character in question, not only to see things from her point of view, but to enter into her mental world.
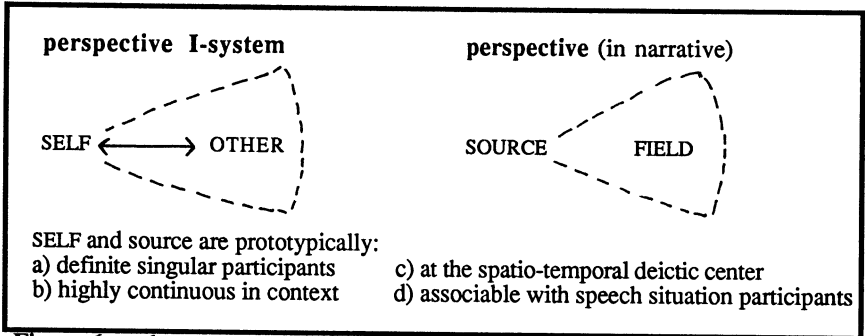
perspective I-system   perspective (in narrative)

SELF ←——————→ OTHER   SOURCE   FIELD

SELF and source are prototypically:
a) definite singular participants   c) at the spatio-temporal deictic center
b) highly continuous in context   d) associable with speech situation participants

Figure 6: analogous properties of the perspective indexical system and the concept of perspective

3.1. Psychological Perspective and Binding Distance.

Figure 7 represents our hypothesis that binding distance correlates with the degree of access to the source character's mind. As access to the character's perceptual and thought processes increases, the more presupposed the character becomes as the source of what is presented in the text. And the more presupposed the character becomes, the less is the need for a proximate antecedent in the text. Note that by the same reasoning 1st person reflexives do not require an antecedent in interactional discourse.

```
<--------------------------------Binding distance--------------------------------->
      free                    long distance                local

I-----------------------I--------------------I--------------------I

   full(a)            more(b)              little(c)           none(d)
              Degree of accessibility to the character's mind
```
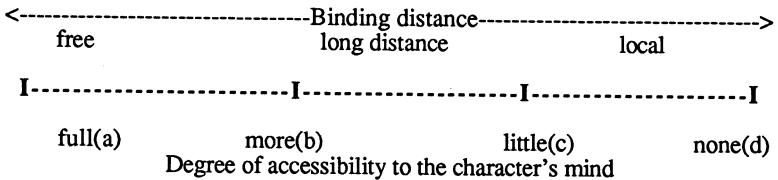
Figure 7: Allowable syntactic distance between SELF and its antecedent, and
its relation to the accessibility of the character's mental world.

Point (a) on the continuum represents the fact that in represented thought context occurrences of SELF are syntactically free, in the sense that they do not require an antecedent in the same sentence. Indeed, SELF may itself be sentence subject, and its antecedent may be many sentences back. Point (b) represents the fact that in psychological reports SELF-forms must find their antecedent within the same complex sentence. Pure spatio-temporal perspective (point c) marginally permits long distance binding if a psychological state of the source character is pragmatically inferred. Finally, in non-perspectival contexts (point d) a SELF-form is locally bound, i.e. it behaves like a garden variety clausal reflexive.

3.2. Free Reflexive.

Example (17) is typical represented thought from narrative discourse. Note the syntactically free occurrences of SELF in lines d and e, which refer to Yu Gentu. These represent Yu Gentu as an object of his own consciousness, a phenomenon known as reflective thought. Note that when the text shifts out of represented thought at line f, a plain pronoun rather than SELF must be used to represent Yu Gentu. It is important that the SELF form stands for the source's psychological reflection on himself, for this motivates the complete lack of antecedent in the sentence, just as first person reflexives can be unbound in interactional discourse.

Example (18) illustrates the same effect in Korean. The syntactically free occurrence of <u>caki</u> in line c is self reflective, and the context is represented thought, indicated e.g. by the rhetorical question and the experiential form of the clause in line e. On the other hand, <u>caki</u> in line a is in a psychological report context characteristic of long distance binding, to be discussed in 3.3, and is not self reflective.

(17) [Yu Gentu, the perspectival source, has discovered an ancient tomb, and the unscrupulous Yu Wanggou is trying to trick him into revealing its location.] '(a) Eyes cast down, Yu Gentu did not answer, but his mind was awhirl: (b) that day, when (he) discovered the old tomb, there had been nobody around! (c) If Yu Wanggou had seen what was happening, wouldn't (he) have come over right there and then? (d) Must be that from far away (he) saw SELF(=YG) digging on the ground, had a suspicion, (e) and now is bluffing SELF(=YG)! This slippery fellow is indeed full of tricks. (f) Thinking like this, (his) mind calmed down somewhat, he lifted up (his) eyes and said, ...'   *Gu Mu* : p.14

a. *Yú Gēntǔ chuí zhe   yǎnpí bù kēngshēng,   xīn lǐ   què zài dǎzhuànzhuan:*
   Yu Gentu hang DUR eyelid not make-sound heart in whereas DUR whirl

b. *nà yītiān,   ∅ fāxiàn   gǔmù   de shíhòu, sìzhōu míngmíng méiyǒu rén na!*
   that one-day, discover old-tomb NL time   around clearly no   person SFP

c. *Yàoshì Yú Wanggǒu kànjiàn le, ∅ dāngchǎng   hái huì bú guòlái?*
   if   Yu Wanggou see   PFV there-and-then still would not come-over

d. *∅ zhǔnshì yuǎnyuǎnde kànjiàn ZIJI zài dìxià   kōuba, ∅ qǐ le yíxīn,*
   must-be from-far-away see   SELF at ground-down dig   raise PFV suspicion

e. *∅ lái   zhà ZIJI li!   Zhè ge huátóu   shǒuduàn shì mǎn duō de.*
   come bluff SELF SFP   this CL slippery-fellow trick is quite many

f. *Zhème yī xiǎng, xīn lǐ shāowēi tāshi le   xiē,   tā tái qǐ yǎn shuō, ..*
   thus once think, heart in slightly steady CRS some   he lift up eye say

(18) [Kyounghun and Kyoungsoo are half-brothers who live with their father. The older Kyounghun is reflecting on the younger Kyoungsu's difficulties with their father.] 'Suddenly Kyounghun felt that SELF was not in a position to either defend Kyoungsoo or to agree with what father said. While SELF was gone, what could Kyoungsoo and father have been thinking, living in the same house with such different perspectives. The desire to smoke overcame Kyounghun.'

a. sunkan Kyounghun-un *CAKI*-ka esaekhakey-nama Kyoungsu-lul onghoha-nun
   suddenly Kyounhun-TOP SELF-NOM awkwardly-even Kyoungsu-ACC defending

b. ipcang-to   mot-toi-mye, kulettako apeci-uy mal-ey tongcohal-su-to
   status-either not-being-and, but-then father-GEN talk-to agree-can-either

c. epum-ul nukki-ess-ta. *CAKI*-ka ep-nun tongan apeci-wa Kyoungsu-nun
   not-ACC felt   SELF-NOM absent period father-and Koungsu-TOP

d. iletkey selo   talun   sisen-ul   kaciko han cipung-mit-eyse
   this-way each-other different perspective-ACC having same roof-under-LOC

e. sal-myense muet-ul saengkakha-ess-ul-kka. Kyounghun-nun tampae-ka piuko sip-ess-ta.
    living   what-ACC   thought-Q   Kyounghun-TOP cigarette-NOM smoke wanted

(19) shows that unbound SELF-forms in represented thought context are necessarily self-reflective. It is a passage of non-reflective represented thought in which the source's emotion is directly portrayed in a way for which there is no matching convention in English. Note that the source is represented by zero anaphora, and by plain pronouns in lines (a) and (b), but not by SELF. If <u>ziji</u> is substituted for <u>ta</u> in lines a and b, the passage becomes self-reflective in the sense that the character is not just experiencing the emotions, but is reflecting on her experience of them.

(19) 'For such a long time she'd been like a thief in the matter of love, like a beggar; pursue, beg, hide -- enough suffering!'   *Ye yu Zhou* : p. 717

a. *Zhème cháng shíjiān yǐlái, tā xiàng shì aìqíng shàng de   zéi,*
    such   long   time   since she as-if is love on ASSOC thief

b. *zhuīláizhuīqù, qiúláiqiúqù, duǒláiduǒqù, tā shòu   gòu le!*
   pursue   beg   hide   she suffer enough CRS!

3.3. Long Distance Binding (LDB).

Next we turn to LDB contexts, in which we claim the listener should only have partial access to the source character's mental world, i.e. the sentence should be an explicit or implicit psychological report. In other words, the antecedent of the SELF-form should be the source of perspective in a psychologial sense (cf Kuroda 1973). As such a position of syntactic prominence should be preferred, but not required, for the antecedent. This reasoning parallels Kuno's study of empathy in syntax, and is reflected in O'Grady's (1987) accessibility hierarchy for Korean, exemplified in (20). In (20a) SELF has three possible antecedents, but prefers the sentence subject. In (20b) <u>caki</u> in Korean cannot be coreferential with first person, and so prefers the direct object over the oblique. But in (20c) the oblique is the only available antecedent[7].

(20) a. *Insu-nun Suni-eykey Chelsu-uy iyaki-lul   CAKI pang-eyse haecu-ess-ta.*
       Insu-TOP Suni-DAT Chelsu-GEN story-ACC SELF room-in   did-give.
       'Insu told Suni a story about Chelsu in SELF's room.' (SELF => Insu)

   b. *na-nun Suni-eykey Chelsu-uy iyaki-lul CAKI pang-eyse haecu-ess-ta.*
       I-TOP Suni-DAT Chelsu-GEN story-ACC SELF room-in   did-give.
       'I told Suni Chelsu's story (about Chelsu) in SELF's room.' (SELF => Suni)

   c. *na-nun Chelsu-uy iyaki-lul CAKI pang-eyse tul-ess-ta.*
       I-TOP Chelsu-GEN story-ACC SELF room-in   heard
       'I heard Chelsu's story (from Chelsu) in SELF's room.' (SELF => Chelsu)

That the accessibility hierarchy is not the whole story is shown by (21), in which the meaning of the postposition controls the extent to which Chelsu is the psychological source, and correspondingly the acceptability of Chelsu as the antecedent to SELF. The locative <u>lopute</u> suggests that Chelsu is the source of information (in Sell's 1987 sense), and therefore of perspective. The circumstantial <u>eytaehan</u> does not allow this interpretation. Furthermore, in comparing (20b) and (20c), note that antecedence to <u>caki</u> in (20c) coerces a source interpretation on the ambiguous genitive. In order for Chelsu to

be accepted as antecedent to <u>caki</u>, the sentence must be interpreted to involve his perspective.

(21) *na-nun {Chelsu-lopute / *-eytaehan } iyaki-lul CAKI pang-eyse (khachi) tulessta.*
   I-TOP {Chelsu-from/ *-about }   story-ACC SELF room-LOC (together) heard.
   'I heard a story from Chelsu / about Chelsu in SELF's room.'

   Example (22) supports the claim that with LDB reflexives the real issue is the extent of access to the character's psychological world, not the character's consciousness (too strong), and not the mere spatio-temporal perspective of the character (too weak). In (22a), a perception verb introduces the content of perception. This is a straightforward psychological report, giving the listener limited access to the character's mind, thus licensing the use of SELF in the final clause. In (22b) there is no perception verb, so that on a literal level the sentence encodes only the source's spatio-temporal perspective. But eye-raising is a conventional preparatory act antecedent to visual perception, automatically triggering the inference that Li Rong actually sees the tide. In (22c) leaning up is a posture change that can promote perception, although it is not taken as a conventional preparatory act as in (b). Thus the listener may infer that Li Rong saw the wave,

(22) *Lǐ Róng zài shātān shàng tiāo le yī kuài dìfāng, tǎng le xiàlái. Shí fēnzhōng hòu,...*
   Li Rong at beach on select PFV one CL spot, lie PFV down ten minutes after
   'Li Rong selected a spot on the beach and lay down. Ten minutes later,...'

   a. *tā tái yǎn cháo qián kànqù,*   a. he raised his eyes to look up,
      he raise eye towards front look-up

   b. *tā tái qǐ yǎn,*   b. he raised his eyes,
      he raise up eye

   c. *?tā qiànqǐ shēnzi*   c. ?he leaned up,
      he raise-up body

   d. **tā chénrù mèngxiāng,*   d. *he fell asleep,
      he sink-in dream

   e. *Tā réng wèi zhùyìdào*   e. He still did not notice,
      he still not notice

   cháoshuǐ zhèng jiànjian xiàng   ZIJI   yǒng lái.
   tide   DUR gradually towards SELF surge come
   the tide was gradually coming towards SELF.'

since he was in a position to do so. Indeed, (c) is acceptable *only* if the listener makes this inference. Note that in any case, (c) is worse than (b), but better than (d); (22d) shows that mere spatio-temporal perspective, where psychological inferences are specifically excluded, is not enough to license a LDB reflexive, contrary to Sell's (1987) position. Thus spatio-temporal perspective seems able to license LDB just in case the character's psychological state is exposed. On the other hand, actual consciousness is not the issue with LDB reflexives, as claimed by Hong (1989) and others. In (22e) the approaching wave is explicitly not part of Li Rong's consciousness. However, the entire sentence is a psychological report concerning his awareness of the tide.

There may appear to be a clear break between free reflexives on the one hand, and LDB reflexives on the other. But we believe there are transitional cases, as in (23a) and (b). (23a) is a standard instance of represented, reflective thought in which the occurrences of SELF are syntactically free.

(23)a. i. *Wáng Líng duìzhe jìngzi dǎliàng zhe   ZIJI   de liányīqún.*
      Wang Ling face   mirror examine DUR· SELF GEN dress

    ii. *ZIJI   de   yāo   hěn xì.   ZIJI zhēn piàoliàng.*
       SELF GEN waist very thin SELF really pretty

    iii. *Búguò ZIJI   yào   zhùyì   búyào chī tài yóunì de dōngxi.*
        but   SELF should be-careful not   eat too fatty NL thing

    '(i) Wang Ling looked in the mirror to check her dress. (ii) SELF was pretty. SELF has small waist. (iii) But SELF should be careful of fatty food.'

In (23b), this self-reflection of Wang Ling's is recast as projected thought: Li Rong is thinking about Wang Ling's opinion of herself, and is making fun of the way she thinks of her looks. Important for us is that the occurrences of SELF take Wang Ling (the projected ego), and not Li Rong (the source) as their antecedent, and yet for the listener only Li Rong's thoughts are directly accessible. In this case we do not get a simple correlation between represented thought and free reflexives.

(23) b. i. *Lǐ Róng hěn shēngqì.   Wáng Líng tài   zìgāozìdà le!*
       Li Rong very irritated   Wang Ling too conceited CRS

    ii. *ZIJI piàoliàng, ZIJI cōngmíng, biérén dōu bǐbúshàng   ZIJI ...*
       SELF pretty   SELF smart   OTHERs all cannot-compare-with SELF

    '(i) Li Rong felt irritated. Wang Ling was so conceited! (ii) SELF was pretty, SELF was smart, no other person was as good as SELF...'

    Occurrences of SELF in LDB contexts have been claimed to be a P-pronominal variant. An important property of P-pronominals is their ability to be non-coreferential, as in (24a). But successive coordinate occurrences of SELF are necessarily coreferential as shown by (24b).

(24) a. *Tā$_i$ yīnggāi gāoxìng cáiduì, yīnwei dàjiā   dōu xǐhuān tā$_{i/j}$.*
      he should   happy   right   because everyone all like him
      'He$_i$ should be happy, because everyone likes him$_{i/j}$.'

    b. *ZIJI yīnggāi gāoxìng cáiduì, yīnwei dàjiā   dōu xǐhuān ZIJI $_{i/*j}$.*
      SELF should happy   right   because everyone all like SELF
      'SELF$_i$ should be happy, because everyone likes SELF$_{i/*j}$.'

This necessary coreference is consonant with the claim that SELF codes the source of a conceptual perspective, so that non-coreference would entail not just shift in referent, but shift in the psychological perspective from which the story world is viewed. Note also that SELF cannot be split into two people in reciprocals, as is the case, for example, with German reflexives. In (25) an overt morpheme for *eachother* must be used, rather than

SELF. Conceptually this means that there cannot be two simultaneous contradictory perspectives coded in the same sentence.8

(25) *Suni-wa Yongi-nun* { *SELO/*CAKI-lul* } *kkyean-ass-ta.*
   Suni-with Yongi-TOP {each other/SELF-ACC} hug.
   'Suni and Yonghi hugged (each other).'

3.4. Local Binding. Where do locally bound reflexives fit in the cognitive framework we have set up for the analysis of SELF-forms? Examples (26) and (27) address this final issue. In (26) the introductory adverbial clauses (a) and (b) provide syntactically equivalent antecedents for the SELF-form in the following clause: (a) is a perspectival introduction, since it sets up Insu as a source; (b) is a non-perspectival introduction: although Insu is mentioned, the speaker does not take his point of view. Note that only (a), the perspectival introduction, provides an acceptable antecedent to the SELF-form. This confirms the previous data showing that the LDB relation must be accompanied by the character's psychological perspective, as claimed by Kuno and others.

(26) *Suni-nun himcue malha-ess-ta.*   'Suni spoke bluntly:
   Suni-TOP firmly spoke.

   a. *"Insu-uy kwancem-eyse po-myeon,*   "From Insu's point of view...
         Insu-GEN point-of-view-from see-if,
   b. ?? *"Insu-eytaehae malhaca-myeon,*   "Speaking of Insu...
         Insu-about saying-if,

   i saken-un *CAKI-hantae khun sonhae-lul cu-ess-ta"*   the accident hurt SELF badly.'"
   this accident-TOP SELF-to big damage-ACC affected.

   Example (27) is exactly equivalent to (26), except that now there is an available antecedent within the clause, and the non-perspectival introduction of (27b) is now acceptable. This suggests that the local-binding relation does not entail the perspective of a source the way the LDB relation does. However, within clauses such as (27b) O'Grady's accessibility hierarchy still seems to hold, suggesting that even in the absence of an actual perspectival source, the grammatical subject is still the preferred antecedent of a SELF-form.

(27) *Suni-nun himcue malha-ess-ta.*   'Suni spoke bluntly:
   Suni-TOP firmly spoke.

   a. *"Insu-uy kwancem-eyse po-myeon,*   "From Insu's point of view...
         Insu-GEN point-of-view-from see-if,

   b. *"Insu-eytaehae malhaca-myeon,*   "Speaking of Insu...
         Insu-about saying-if,

   ku-nun CAKI-hantae khun sonhae-lul cu-ess-ta"   he hurt SELF badly.'"
   he-TOP SELF-to   big   damage-ACC gave.

This suggests that Kuno's intuition about the empathy effect with locally-bound reflexives is correct. They seem to be a syntactic fossilization of the perspective phenomenon in discourse.

4. Conclusion. Unlike syntax-based analyses which consider clause-bound reflexives as a starting point and LDB reflexives as an extension of this syntactic environment, but

351

which have little to say about free reflexives, we have presented an analysis which takes syntactically free occurences of reflexive morphemes in contrastive and represented thought contexts as a starting point and extended the analysis into the domain of LD and local binding. This approach has the advantage of showing how apparently unrelated uses of reflexives can be derived from a unitary source: the perspectival indexical system. On the other hand, it may not be able to account for all the syntactic details of long-distance binding, and has little to say about local binding of reflexives, although it is not in conflict with it. We thus regard the syntactic and the discourse-semantic approaches to the problem as complementary rather than antagonistic.

NOTES

* This research was supported in part by National Science Foundation Grant No. IRI-8610517. We would like to express our gratitude to members of the Deictic Center Research Project of the Center for Cognitive Science at SUNY, Buffalo for their comments and discussions. Special thanks go to Matthew Dryer, Mary Galbraith, George Lakoff, Joongwon Park, Jackson T.-S. Sun, Randy J. LaPolla, Dan Devitt, Fengxiang Li, Eve Sweetser and Dingxu Shi for their comments on the earlier version of this paper.

1 See Reinhart (1983) for the standard binding analysis; Tang (1989) and Cole et al. (1989) for a syntactic approach and Kuroda (1973), Kuno (1987) and Sells (1987) for conceptually-based approaches to long distance binding; and Zribi-Herz (1989) for unbound reflexives. The latter provides a current overview of issues. Although these assumptions are characteristic of the literature, not all of them are consistently held. For example, Kuno (1987 and earlier) has consistently argued that empathy licenses long distance reflexives; and Zribi-Herz argues that unbound occurrences of SELF-forms in represented thought are an autonomous phenomenon based on "subject of consciousness." Van Hoek(1990) provides a unified cognitive semantic account of reflexives in English equivalent in spirit to our treatment of Korean and Mandarin reflexives.

2 In this paper we alternate data from Korean and Mandarin because the two languages show parallel behavior with respect to most of the phenomena discussed. Mandarin does not allow non-subject controllers to the extent that Korean does (see ex 20). Korean has an additional reflexive morpheme casin which occurs in 1st and 2nd person clause-bound contexts, to the exclusion of caki. Our data show, however, that caki does occur in 3rd person clause-bound contexts, and casin in free contexts, so that the claim that casin is a bound and caki a free pronoun is too simplistic (cf Cole et al. 1990). We are aware of the caki-casin problem, but do not address it here.

3 These judgements pertain to a non-referential reading. A specific referent for OTHER can be placed at the deictic center, e.g. if OTHER is taken to be the addressee.

4 George Lakoff (personal communication) points out that the radial category in figure 4 can be collapsed in the context of a frame semantic analysis. For example, the volitional inferences at the bottom of the table are only illustrative of a wide range of variations produced by the imposition of very specific situation frames.

5 Korean does not allow this specific inference, but has a separate morpheme for it cecelo.

6 Korean psych-verbs are much more likely to have an unexpressed (zero) experiencer in the experiential construction (represented thought) than they are in the agentive construction (psychological report)--see Chun & Zubin, 1990 in this volume.

7 In Madarin only the subject is accessible as antecedent in (20).

8 We have found complex cases of syntactically adjacent non-coreferential occurrences of SELF which involve projected thought, that is, the embedding of one character's perspective in the reflective consciousness of another character, as in (23b).

REFERENCES

BANFIELD, A. 1982. Unspeakable sentences. Boston :Routledge and Kegan Paul.

CHO, YOUNGMEE YU. 1985. An LFG analysis of the Korean reflexive caki. Proceedings of Harvard Linguistics Workshop.

CHUN, SOON AE AND ZUBIN, DAVID A. 1990. Experiencial vs. agentive constructions in Korean narrative. In this volume.

CLEMENTS, GEORGE N. 1975. The logophoric pronoun in Ewe: its role in discourse. Journal of West African Languages 10.141-77.

COLE, P., G. HERMON and L.-M. SUNG. 1990. Principles and Parameters of Long-Distance Reflexives. Linguistic Inquiry 21.1-22.

HONG, KI-SUN. 1989. Consciousness condition on the Korean reflexive caki. Berkeley Linguistics Society 15.

KURODA, S.-Y.1973. Where epistemology, style and grammear meet: A case study from Japanese. A Festschrift for Morris Halle, eds. by S. Anderson and P. Kiparsky, 377-391. New York: Holt, Rinehart & Winston.

KURODA, S.-Y. 1976. Reflections on the foundations of narrative theory. Pragmatics of language and Literature, ed. by T.A. Van Dijk. Amsterdam: North-Holland Publishing Company.

KUNO, SUSUMU. 1987. Functional syntax. Chicago: University of Chicago Press.

KUNO, SUSUMU.1973. The Structure of the Japanese language. Cambridge, Mass.: MIT Press.

LI, RENJIAN. 1984. On ziji and constructions with ziji. Zhongguo Yuwen 179.102-7.

O'GRADY, WILLIAM. 1987. The interpretation of Korean anaphor. Language 63.251-77.

RAPAPORT, WILLIAM J., et al. 1989. Deictic centers and the cognitive structures of narrative comprehension. Tech Report 89-01: Dept. of Computer Science, University at Buffalo.

REINHART, TANYA. 1983. Anaphora and semantic interpretation. Chicago: U. Chicago Press.

SELLS, PETER 1987. Aspects of logophoricity. Linguistic Inquiry 18.445-79.

SUN, JACKSON T.-S. 1989. Toward explaining reflexive constructions in Mandarin Chinese. ms. Berkeley.

TANG, CHIH-CHEN J. 1989. Chinese reflexives. Narural Language & Linguistic Theory 7.193-21.VAN HOEK, KAREN. 1990. Dissertation in progress. UCSD.

WIEBE, JANYCE. 1989. Recognizing subjective sentences, Ph.D. diss. U. at Buffalo.

ZRIBI-HERTZ, ANNE. 1989. Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. Language 65.695-727.

TEXT REFERENCES

HAN, SOOSAN. 1980. Haepingki-uy achim. vol 2. Seoul.

KE, YUNLU. 1986. Ye yu Zhou. Renmin Wenxue. Beijing.

XU XIAOYU, 1982. Gu Mu. Shouhuo, no. 5. Beijing.

PARASESSION

ON

THE LEGACY

OF

GRICE

THE VIOLATION OF GRICE'S
MAXIMS IN JOKES

Salvatore Attardo
Purdue University

The principal claim of this paper (1) is that the study of jokes and other kinds of humorous texts can yield interesting insights on the nature of cooperative linguistic communication and, more specifically, on the relative status of the maxims. It is also claimed that these insights would be difficult to achieve without taking humor research into account.

A basic assumption which underlies the following remarks is that a large number of jokes involve violations of one or more of Grice's maxims. Consider the following examples:

QUANTITY (1) 'Excuse me, do you know what time it is?' 'Yes.'

RELATION (2) 'How many surrealists does it take to screw in a lightbulb?' 'Fish!'

MANNER (3) 'Do you believe in clubs for young men?' 'Only when kindness fails.' (Attributed to W.C. Fields)

QUALITY (4) 'Why did the Vice President fly to Panama?' 'Because the fighting is over.' (Johnny Carson 1-19-90)

Example (1) violates the maxim of quantity by not providing enough information. Violation through providing excess information is also possible and, for instance, was codified in medieval French literature under the form of 'enumération' (e.g., Garapon 1957:22-25). Example (2) is an 'absurd' joke, with a certain twisted appropriateness from the well known surrealist taste for bizarre associations. Example (3) violates the 'submaxim' of manner 'avoid ambiguity,' as in general do all forms of verbal humor based on ambiguity, such as puns. Example (4) is a deliberate infraction of the maxim of quality.

It should be noted that what is being claimed is that the above texts do not flout or exploit the maxims, but that they violate them, i.e., they fail to conform to their 'recommendations' and thus constitute examples of non-cooperative behaviour. Nevertheless, the examples do 'somehow' make sense, and are understood and recognized as jokes. The solution to this apparent puzzle will be presented in the first part of the paper. But first some necessary observations on the processing of the joke texts will be introduced, as a precondition to the explanation.

Grice notes that by violating one of the maxims the speaker 'will be liable to mislead' (1975:49); and this is exactly the case in the text of a joke in a literal processing. The processing of a joke can be described (in theory-neutral terms) as the discovery of a second 'sense' in a text that had initially seemed to be headed in the direction of a 'normal' disambiguation. The script theory proposed by Raskin (1985) describes this phenomenon as the imposition of a second 'script'; the structuralist-based theories, summarized in Attardo (1989) as the discovery of a second 'isotopy'

(cf. Greimas 1966 and references in Attardo 1988, 1989). The violation of the maxims is related to the 'unexpected' presence of the second sense (script, isotopy). The speaker producing the text uses the violation of a maxim to mislead the hearer into believing that 'normal' reliable information is being provided, while in effect the text, or the utterance, is rigged (cf. Morreall 1983:79-82; Hunter 1983).

The process of 'getting a joke' can be roughly schematized as the processing of the text, establishing 'on line' its sense (first script/isotopy), until an element is encountered (the punch-line) that defeats that sense and forces a backtracking to the beginning of the text (or to some other significant point) and a reinterpretation (second script/isotopy) of the processed text.

Consider example (3) again. The polysemous word 'club' introduces a first (unnoticed) ambiguity, while 'kindness fails' is the part of the sentence which redirects the interpretation of the text onto the second script/isotopy. During the process of disambiguation the reader selects the 'social activities' meaning of the polysemous lexeme 'club.' Nothing up to the point when the VP 'fails' forces the reader to correct his/her choice. When the word 'fails' is reached, however, the disambiguation process is brought to a halt. It is impossible, at that point, to make sense of the sentence. The reader is then faced with an option: either discard the text as ill-formed, and thereby assume that the utterance did not convey any meaning (besides its own ill-formedness, and the inferences thereof), or backtrack and check for possible ambiguous or polysemous lexemes, constructions, etc., that might be given another reading. The option of declaring the text ill-formed is undesirable under the principle of cooperation (see below). While backtracking, the reader encounters once again the lexeme 'clubs.' The sense 'stick' offers itself to the reader, who can then reprocess the second part of the text as a (very) elliptic sentence having roughly the form (I would use clubs on young men) 'only when kindness fails.'

It should be pointed out, for the sake of precision, that the disambiguation process described above is only one of the possible configurations of the 'punchline' found in jokes. The punchline may not involve an actual backtracking (i.e., the ambiguity and the element that forces the reevaluation may coincide), or may be 'scattered' along the text (for example, in jokes based on alliteration) (cf Attardo et al. in print for further details).

Some of the issues raised by this apparently non-cooperative behaviour are as follows: 1) the nature of the communicative status of humorous texts, 2) the implicit/explicit balance, 3) the relative status of the maxims.

1) The communicative status of humorous texts.

The first step towards the solution of the apparent puzzle of the understanding of non-cooperative texts, such as jokes, will be looking at an alternative set of maxims, proposed to account for the 'non-cooperative' behaviour of jokes. Next, attention will be given to socially accepted activities performed 'using' jokes. Finally, we will consider the status of the communicative mode of jokes.

If humorous texts violate the maxims, one would expect them to become non-cooperative and/or to lose meaningfulness. Nevertheless, jokes are 'understood,' and are not perceived as 'lies' (lying is non-cooperative), or as ill-formed, or cryptic

texts. To account for this fact, Raskin (1985) suggested that joking involves a different kind of 'communication mode,' governed by a different set of maxims. The maxims for that 'non-bona-fide' mode are the following:

Quantity : Give exactly as much information as is necessary for the joke

Quality : Say only what is compatible with the world of the joke

Relation : Say only what is relevant to the joke

Manner : Tell the joke efficiently (Raskin 1985:103)

Thus, the apparent paradox is solved: the hearer will backtrack, after realizing he/she has been misled, and will reinterpret the information provided in the text on the basis of the 'humor' maxims, switch to the non-bona-fide mode of humor, and react accordingly (i.e., laughing, smiling, etc.). It is necessary, then, to distinguish a first reading of the joke, in which the reader notices the violation of Grice's maxims, and a second reading, in which the reader, having switched to the non-bona-fide mode of humor, reinterprets the text as a joke, and thus accepts strange and un-realistic events ('suspension of disbelief'), activates particular stereotypes, and in general 'tunes in' to the idiosyncrasies of the non-bona-fide mode of humor.

What is then the communicative status of jokes? As it has been shown, jokes involve the violation of one (or more) maxims in the first reading. Jokes have, however, been shown to perform various communicative functions. Drew (1987) analyzed reactions to humorous teasing and found that many speakers take teasing seriously, at face value, thus clearly showing that they assume that the utterer of the tease is communicating effectively. Zhao (1988) has shown that jokes can convey relevant 'bona-fide' information as, for example, in the case of jokes about an unfa-miliar situation/culture. Mulkay (1988) discusses several 'uses' of joking (including sociological accounts of the use of humor among the members of a staff hospital, and in a restaurant, as a method of 'picking up' members of the opposite sex); he concludes that by using humorous utterances, the speakers can avoid committing themselves too strongly to what they say. Jefferson (1984) analyzes narratives re-lating problematic situations and finds that in their narratives speakers intersperse humorous remarks to show that 'they can take it.'

It has been noted (Raskin 1985) that after a hearer experiences an apparent failure to reconcile utterances with his/her own belief system, he/she engages the default communicative mode of 'joking.' If the speaker is faced with an utterance whose contents he/she cannot reconcile with his/her knowledge of the world, the speaker will try to assimilate it, either by including the new information in his/her world representation, or by refusing the conflicting information status of 'reliable' knowledge. The joking mode ('Are you kidding?') seems to be the first option, which reflects the premise that joking is more socially acceptable than lying or not making sense (cf. Raskin 1985:104). Raskin thus hypothesized an 'extended form of bona fide communication' incorporating humor (and governed by both Grice's maxims and the 'humor maxims'). From the foregoing discussion, it should be clear that 1) speakers use humorous texts cooperatively (thus corroborating Raskin's thesis),

but also 2) they rely on the 'subversion' of the maxims to achieve socially desirable effects. Consider, for instance, the possibility of 'backing out' of an utterance, by claiming that one 'did not really mean it' (i.e., that one was infringing the quality maxim).

It seems that a radical dichotomy between 'serious' bona-fide use of language, and 'humorous' non-bona-fide cannot be maintained. Grice's hypothesized speaker, totally committed to the truth and relevance of his/her utterances, is a useful abstraction, but should be considered only as such. In reality, speakers engaged in everyday communication use humorous remarks that the hearers decode, interpret as such, and use along with other information to build their vision of the communicative context.

The consequences of this recognition–that communication which violates the maxims can still be 'cooperative'–are far ranging. Any attempt to characterize linguistic interaction will have to incorporate rules and inferential mechanisms to handle humorous violations of the principle of cooperation.

Now that we have assessed the communicative status of jokes and other humorous types of texts, it is possible to consider more carefully the 'implicit' dimension of jokes.

2) The importance of the implicit in jokes.

It has been frequently noted that some part of the information must be left implicit in jokes. Explication of the mechanisms involved in the humorous effect of the text results in the destruction of the humorous effect: i.e., a joke loses its humor when the joketeller explains the punchline. Eco (1981), after claiming that all jokes involve, among other mechanisms, the violation of a 'rule,' notes that the rule must be left implicit. Dolitsky (1991) mentions that the way in which the information is organized is relevant to the 'structure' of the joke–that is, not every formulation of the information contained in the joke text (and inferrable from it) will be considered a successful joke. It is precisely because part of the information is present only in the implicit part of the text that the joke acquires one of its characteristics. In other words, for the joke to 'function' as such, some information must be left unsaid: i.e., Grice's maxim of quantity must be violated.

The modality of this delicate explicit/implicit equilibrium has yet to be explored fully. A few preliminary remarks will serve the purpose of delimiting the range of the problems involved.

It has been noted that the resolution of incongruity in humor involves mental expenditure (cf. for instance Freud 1905). This fact alone would require that the hearer of a joke be required to infer some implicit information, or to perform some cognitive task.

Another well known requirement of the 'punchline' of a joke is that it should come 'unexpectedly' (this is commonly referred to as the 'surprise' theory of humor). Once one takes into consideration such notions as surprise, or expectedness, it becomes necessary to refer to the linear aspect of the text of the joke (for a discussion cf. Attardo 1989). In short, as we have briefly mentioned, the decoding of the text of the joke is a temporally structured activity in which the various elements are

necessarily introduced in a linear order. This implies the necessity of avoiding the introduction of the 'second script' in a text engaged in actualizing the first one, i.e., giving away the punchline early, thus violating the need for surprise. This fact seems to account for the often noted but scarcely explored fact that the punch-line of the joke comes towards the end of the text. In fact it has been claimed (Oring 1989, Attardo 1989, Attardo et al. in press) that the punch-line must occur finally in the text, and that the exceptions can be predicted fairly accurately.

The requirement that the presence of the second sense not be introduced early in the text applies not only to explicit mentions of elements of the second script, but also to any related element which could enable the hearer's actualization of the script via inferential channels. This is clearly connected to the concept of 'manifestness' introduced by Sperber and Wilson (1986) for all the contextual information which can be brought into the focus of the speaker's attention. In this terminology, the text of the joke must render non-manifest the presence (or the future presence) of an alternative script.

Consider the following example:

(5) A young lady was talking to the doctor who had operated upon her. 'Do you think the scar will show?' she asked. 'That will be entirely up to you,' he said.

The joke rotates on the passage from the 'medical' script to the 'nudity' script. The allusion to nudity cannot be topicalized before the end of the text otherwise the joke would lose its effectiveness. If the first sentence were substituted by 'A young nudist lady...' the punchline would not only lose its suddenness, but would probably lose its evocative side (nudity– weakly–implies sex).

The quantity maxim for jokes: 'Give as much information as is necessary for the joke' can now be viewed as an informal algorithm for the computation of the quantity of information to be left implicit.

3) Relative position of the maxims.

In this section, evidence for a hierachical organization of the maxims will be presented, first from an empirical analysis, and then on theoretical grounds.

In an analysis of some 243 jokes extracted from a corpus of 6500, Van Raemdonck (1986, 1989) found that all the jokes violated the maxim of relevance, while only some violated another maxim as well (Van Raemdonck 1986:62-63). Furthermore, the violations were interdependent. Although the figures are not claimed to be statistically reliable, they still retain interest as a well grounded example.

This seems to suggest that when any of the other three maxims is violated in a joke, the maxim of relevance is necessarily violated as well. If the speaker does not believe in the truth of what he/she is saying, the content of the utterance can hardly be expected to be relevant (though the speaker could be lying, thus producing a relevant but not-cooperative utterance; however, this would not qualify as a joke). If the speaker does not provide enough information (or provides too much information), he/she will not be relevant, either because his/her information will fail to cover some of the relevant issues, or because the information will cover issues which are

not relevant. If the speaker is obscure or ambiguous, his/her contribution will not be relevant since the hearer will not be able to evaluate whether the information provided is 'to the point.' Thus, it seems that the maxim of relevance subsumes the other three, in the sense that in order to be relevant, one must first be sincere, orderly, and exhaustive.

It should be recalled now that the 'obligatory violation of the maxim of quantity' was individuated as the underlying motivation for the presence of implicit information in the text of a joke. If all jokes must abide by the non-bona-fide quantity maxim (i.e., must violate by not giving enough information Grice's maxim of quantity), there seems to be evidence for a maxim of quantity at the same level of the super-maxim of relevance. The speaker is required, per Grice's maxims, to provide 'enough' information for the text to be processed without problematic falls into ambiguity (cf. (3)). Similarly, the speaker is supposed to provide collateral information that would prevent the sudden introduction of an unexpected second sense or, in other words, to set communication on a 'safe' base of information which will clearly delimit the 'topic' of the interaction, and thus prevent a premature switch in the topic of a text like (4), where the topic switches from politics to a 'Dan Quayle' slur.

We are thus faced with two claims for 'underlying' maxims: relation and quantity. It is interesting to note that both positions have been claimed by independent research. Sperber and Wilson (1986) propose an underlying super maxim of relevance, while Horn (1984) proposes two 'principles' 'Q' and 'R' to 'evoke,' á la Chomsky, Grice's maxims of quantity and relevance. This is not the place to go into the details of an evaluation of both proposals, but it may be noted that, since both quantity and relevance have been noted to be necessarily infringed upon in a joke, Horn's dualism seems better supported by the facts.

This necessarily brief discussion suggests that the violation of maxims in jokes provides an independently motivated external confirmation to the so-called 'relevance' theory (Sperber and Wilson 1986) and to the 'revised' maxims proposed by Horn (1984), both of which grant to the maxim of relevance (and of quantity) a higher status than the original Gricean text.

Conclusion.

The cooperative aspects of humor as a non-bona-fide mode of communication have been stressed, as well as the need for a revision of the idealized 'bona-fide' mode of communication. It has also been shown that jokes and other kinds of humorous texts can yield information, both on the principled construction of texts which violate the maxims to exploit the deception of the hearer's expectations, and on the hierarchical organization of the maxims.

NOTES

(1) The author wishes to express his gratitude to Victor Raskin, Monica Macaulay, Donalee Hughes and Dan Van Raemdonck. All imperfections, however, should be attributed to the author.

REFERENCES

Attardo, Salvatore. 1988. Trends in European Humor Research: Towards a Text Model. *HUMOR: International Journal of Humor Research.* 1:4. 349-369.

—. **1989.** Morfologia della barzelletta. *Studi Italiani di Linguistica Teorica ed Applicata.* 3. 1989.

Attardo, Salvatore, Paul Baltes, Donalee Hughes, and Marnie J. Petray. (in print) The Linear Organization of Jokes: Statistical Analysis of Two Thousand Texts. *HUMOR: International Journal of Humor Research.*

Dolitsky, Marlene. 1991. Aspects of the Unsaid in Humor. *HUMOR: International Journal of Humor Research.* 4:4. 1991.

Drew, Paul. 1987. Po-faced Receipts of Teases. *Linguistics.* 25. 219-253.

Eco, Umberto. 1981. Il comico e la regola. *Alfabeta.* 21. 5-6. In *Sette anni di desiderio.* Milano: Bompiani. 1983. 253-260. Tr. *Travels in Hyperreality.* New York: Harcourt Brace, 1986.

Freud, Sigmund. 1905. *Der Witz und seine Beziehung zum Unbewussten*

Garapon, Robert. 1957. *La fantaisie verbale et le comique dans le théatre français du moyen-age à la fin du XVIIIème siecle.* Paris: A. Colin.

Greimas, Algirdas Julien. 1966. *Sémantique structurale.* Paris, Larousse. Tr. *Structural Semantics.* Lincoln: University of Nebraska Press. 1983.

Grice, H. P. 1975. Logic and Conversation. In Cole, P. and J. Morgan (eds.) *Syntax and Semantics.* Vol. 3. *Speech Acts.* New York: Academic Press. 41-59.

Horn, Lawrence. 1984. Toward a New Taxonomy for Pragmatic Inference: Q-Based and R-Based Implicature. In Schiffrin, D. (ed.) *Meaning, Form, and Use in Context: Linguistic Applications.* Washington: George-town University Press. 11-42.

Hunter, Lynne. 1983. On Misapplying the Maxims: a Gricean Look at Wit. *Papers from the Regional Meeting, Chicago Linguistic Society.* Chicago: Chicago Linguistic Society. 19. 195-204.

Jefferson, Gail. 1984. On the Organization of Laughter in Talk about Troubles. In Atkinson, J. M. and J. Heritage (eds.) *Structures of Social Action.* Cambridge: Cambridge University Press. 346-369.

Morreall, John. 1983 *Taking Laughter Seriously.* Albany, NY: State University of New York Press.

Mulkay, Michael. 1988. *On Humor. Its Nature and its Place in Modern Society.* Cambridge: Basil Blackwell.

Oring, Elliott. 1989. Between Jokes and Tales: on the Nature of Punch Lines. *HUMOR: International Journal of Humor Research.* 2:4. 349-364.

Raskin, Victor. 1985. *Semantic Mechanisms of Humor.* Dordrecht: Reidel.

Sperber, Dan and Deirdre Wilson. 1986. *Relevance.* Cambridge, MA: Harvard University Press.

Van Raemdonck, Dan. 1986. Laughstory. Du traitement du principe de cooperation et des maximes de conversation dans les histoires drôles. Doctoral Dissertation. Free University of Brussels (Belgium).

— 1989. Laughstory. Modèle psycho-pragmatique d'explication du comique dans les histoires drôles. Manuscript.

Zhao, Yan. 1988. The Information Conveying Aspect of Jokes. *HUMOR: International Journal of Humor Research* 1:3. 279-298.

Constraints on Interpretation
Diane Blakemore
Southampton University

Some time ago after presenting a detailed analysis of *after all* at a conference I was asked by a member of the audience why I had devoted so much time to such a little word. In fact, *after all* is not that little, and semanticists spend a great deal of time on much smaller words (*all* and *not* for example). But perhaps what he really meant was little in importance. What does a word have to do to be important? Contribute to truth conditions perhaps. It is certainly true that according to this criterion, *after all* could not be considered important. Although the use of *after all* in (1) indicates that his braveness is to be expected given his nationality, we would not want to say that the utterance is false should this not be the case.

(1)   He is brave: he is, after all an Englishman

If contributing to truth conditions is a criterion of a word's importance (or size), then there are quite a few unimportant (or little) words. As many will have recognized, the example in (1) is adapted from one given by Grice. The original is given in (2).

(2)   He is an Englishman: he is, therefore, brave

The fact that Grice said relatively little about *therefore* might be taken to mean that he too thought this was a little word. For, essentially, all he said was that this was an example of linguistic meaning which contributes not to truth conditions but to implicatures. That is, it is an example of a conventional implicature. What I hope to show here is that *therefore* and *after all* do in fact play a very important role in communication, a role which can only be appreciated given a proper understanding of the principles governing the interaction of linguistic meaning and contextual information in the interpretation of utterances. In my book *Semantic Constraints on Relevance* I was concerned with this role insomuch as it seemed to provide an explanation for non-truth-conditional linguistic meaning. However, as I shall show in the latter part of this paper, this role is not confined to those expressions which do not have truth-conditional (or conceptual) content.

Karttunen (1974) and Karttunen and Peters (1975) have linked Grice's notion of conventional implicature to the class of phenomena referred to in the presupposition literature as pragmatic presuppositions (cf Stalnaker 1974, 1975). While they recognize that it is unlikely that everything that has been called a pragmatic presupposition is in fact a case of conventional implicature, they claim that what Grice said about *therefore* applies equally well to such little words as *even, yet* and *too*. *Too*, says Karttunen is a rhetorical device whose presence or absence does not have any bearing on the proposition the sentence containing it expresses, but rather relates the sentence to a particular kind of conversational context. In other words, such expressions impose constraints on the context in which the

utterances containing them must be interpreted. If the appropriate context is not available, then the utterance is inappropriate.

But why should there be such words? We all know that the Gricean explanation for conversational implicature lies in the fact that the act of communicating creates certain expectations - expectations of truthfulness, informativeness, relevance and so on. One might think that conventional implicatures, because they are conventional, should have a very different explanation. However, I am going to argue that the existence of expressions that impose constraints on contexts follows from the very nature of communication, and in particular, from the role played by the assumption that the speaker has tried to be relevant.

It will be remembered that in his discussion of the rationale for the conversational maxims Grice toyed between a social explanation and a cognitive one. He ended up dismissing the idea that the maxims had their origin in the nature of society or culture on the grounds that this did not provide a sufficiently general explanation. However, he warned us that the key to a general, psychological explanation lay in the notion of relevance, a notion that he himself left unexplained.

Sperber and Wilson's (1986) Relevance Theory might be regarded as a response to the challenge implicit in Grice's warning. They argue that an account of utterance interpretation must be based on a general cognitive theory of information processing. The basic idea is that in processing information people generally aim to bring about the greatest improvement to their overall representation of the world for the least cost in processing. That is, they try to balance costs and rewards. Obviously, not every addition of information will count as an improvement. A hearer's representation of the world will not necessarily be improved by the addition of information that it already contains. Nor will it be improved by the presentation of information that is unrelated to any of the information that it already contains. The hearer's aim is to integrate new information with information that is already accessible to her, or, in other words, to recover information that is relevant to her. The point is that in every case her search for relevance leads her to process new information in a context of assumptions supplied either from memory, through the interpretation of the previous utterance or via her perceptual abilities.

In this theory computing the effect of a newly presented proposition crucially involves inference. That is, the role of contextual assumptions is to combine with the content of the utterance as premises in an argument. There is no space here to outline the nature of the inferential abilities that Sperber and Wilson believe to be involved in utterance interpretation. However, it is important to recognize that in their theory propositions are treated not just as logical objects, but as psychological representations, and that inferences are psychological computations performed over those representations. Their basic claim is that assumptions about the world come with varying degrees of strength, and that logical computations assign strength to conclusions on the basis of the strength of the premises used in deriving them. Clearly, any new assumption derived as a conclusion (or contextual implication) from a newly presented proposition in a context of existing assumptions will count as an improvement. However, in this framework a proposition whose presentation enables a hearer to derive an assumption which she has already represented may be relevant in virtue of strengthening the hearer's conviction that it is true: having two

independent pieces of evidence for an assumption will lead a hearer to assign it a greater degree of strength than she would have assigned it on the basis of just one of these pieces of evidence. Finally, because an inference system can be used to test for inconsistencies in the propositions presented to it, a proposition whose presentation leads the hearer to derive a conclusion that is inconsistent with an assumption she holds already may be relevant by virtue of leading her to abandon it.

Since the impact of an utterance depends on the context, the hearer must have some principle by which she chooses the particular contextual assumptions she brings to bear. For logically speaking, any of her beliefs and assumptions may be brought to bear on the interpretation of an utterance, which means that logically speaking, her interpretation isn't constrained at all. Since successful communication does occur and hearers often interpret utterances in the way they are expected to, the hearer's choice of context must be one that can be exploited and manipulated by the speaker.

According to Sperber and Wilson, what the speaker manipulates is the hearer's search for relevance. A hearer will only pay attention to a phenomenon if she thinks it is going to be worth her while. This means that there is no point in you attracting my attention, for example, by speaking to me, unless you think you have information that is relevant to me and hence worth processing. So if you deliberately attract my attention, and if I recognize that you are deliberately attracting my attention, then I will expect that I can recover some contextual effect from your behaviour (utterance). However, as I have said, the hearer is not just interested in obtaining some reward: her aim is to recover the greatest contextual effect for the available processing effort. This means that it is in her interest that the information presented to her is the most relevant information that could have been conveyed by the communicator. But of course, the communicator will have his own aims, and these may lead him to give the hearer information whose impact is less than other information he could have given. The point is that to be worth the hearer's attention it must have some impact.

Intuitively, it is clear that the greater the impact a proposition has on the hearer's representation of the world the greater its relevance. On the other hand, accessing contextual assumptions and using them to derive contextual effects involves a cost, and the cost of deriving them in a small easily accessible context will be less than the cost of obtaining them in a larger, less accessible context. This means that it is in the interests of a hearer who is searching for relevance that the speaker should produce an utterance whose interpretation calls for less processing effort than any other utterance he could have made to achieve the same effects. There is always a number of ways of conveying the same information. However, these may require varying amounts of effort from the hearer. Consider, for example, the discourse sequences in (3) and (4) (both adapted from Blass 1986).

(3)   He went into MacDonalds. The quarter pounder sounded good and he ordered it.

(4)   The river had been dry for a long time. Everyone attended the funeral.

Blass assumes that whereas (3) presents no difficulty for a westerner, the sequence

in (4) will seem incomprehensible. While the mention of *MacDonalds* gives a western hearer access to contextual information that she can use in the interpretation of the rest of the sequence, in (4) there is nothing in the interpretation of the first segment that can be used in the interpretation of the second. In contrast, for a speaker of Sissala (a Niger-Congo language) the sequence in (4) would present no problem, while the one in (3) would seem incomprehensible. A western speaker who wanted to communicate the information in (3) to a Sissala speaker would have to make the required contextual assumptions explicit - for example, by producing the sequence in (5), while a Sissala speaker who wished to communicate with a western speaker would have to produce the sequence in (6).

(5)   He went to a place where food is prepared and cooked. There he saw ground meat which was formed into patties, fried and placed between two pieces of bread ....

(6)   If a river has been dry for a long time, then a river spirit has died. When a spirit dies there is a funeral. The river had been dry for a long time ....

The additional complexity of these utterances is the price the speaker has to pay for successful communication. Notice, however, that the sequence in (5) would be far less relevant to a western hearer than the one in (3). The extra processing is not rewarded by an increase in contextual effects. The same point can, of course, be made about the Sissala example.

But of course no speaker who wished to communicate with a western hearer would produce the sequence in (5). Given that he does want to communicate, it is in his interests to make the utterance as easily understood as possible. Indeed, a western hearer presented with (5) might doubt that genuine communication was intended and refuse to make any processing effort at all. In other words, the presumption of relevance carried by every act of communication has two aspects: on the one hand, it carries a presumption of adequate effect, while on the other, it carries a presumption of minimally necessary effort. Taken together these presumptions define a level of optimal relevance. Sperber and Wilson call the principle which gives rise to the presumption of optimal relevance the principle of relevance.

Notice that according to this framework, the responsibility for success in communication is not shared, but taken by the speaker. The hearer simply goes ahead and recovers the interpretation that is consistent with the principle of relevance. This means that a speaker who has a specific interpretation in mind and wishes communication to succeed must have grounds for thinking that the hearer has immediate access to a context which enables her to recover the right (that is, the intended) interpretation. Clearly, these grounds may be mistaken - communication, as we all know, sometimes fails. But in many cases the speaker has good grounds for thinking that the hearer will supply the correct contextual assumptions and recover the intended interpretation. So for example, if at the end of my allotted time the chairperson produces the utterance in (7), s/he is assuming that I have immediate access to the assumptions that allow me to derive the contextual implication in (8):

(7)   Your time is up
(8)   I must stop talking now

In this case the speaker has grounds for thinking that I would recover the intended interpretation on my own accord. There was no need for him to make it clear that he expected this interpretation to be recovered. That is, there was no need to constrain my interpretation in any way. However, a speaker who has a specific interpretation in mind and has grounds for thinking that the hearer cannot be trusted to supply the required contextual assumptions may direct her towards that interpretation by making a certain set of contextual assumptions immediately accessible thus ensuring their selection under the principle of relevance. There are non-linguistic means of doing this - indirect answers are an example. However, we can now see why there are expressions and structures whose sole function is to guide the hearer in the interpretation of the utterances that contain them. Their use ensures correct context selection at minimal processing cost.

    This is not the place for a detailed examination of such devices. In my book (Blakemore 1987) I was mainly concerned with the analysis of so-called discourse connectives, and I will use these now simply to illustrate in an intuitive way the general point I have just been making. Consider the sequence in (9) which I have presented as part of a dialogue. Although A's utterance provides the hearer with an immediately accessible context for the interpretation of B's, it is not clear (out of context) exactly where the relevance of B's remark lies. It could be relevant as evidence for A's claim; it could be relevant as a specification of the implications (and hence relevance) of A's remark; it could be construed as an explanation for the state of affairs A has described; it could be relevant as something that contrasts with this state of affairs; or it could be intended as an attempt to dismiss A's remark as irrelevant. In real conversation the connection between the two remarks would not have been left unspecified, and B would have made his intentions clear either through intonation alone or through intonation combined with the use of the expressions that preface the responses in (10).

(9)   A: Susan's not coming today
       B: Tom's in town
(10)(a)   After all, Tom's in town
    (b)   So Tom's in town
    (c)   You see, Tom's in town
    (d)   However, Tom's in town
    (e)   Anyway, Tom's in town

    Notice that in order to establish the prescribed connection, the hearer must supply certain contextual assumptions. For example, B's remark can be construed as evidence for A's remark only given the contextual assumption in (11).

(11)   If Tom is in town, then Susan won't be coming

The same point can be made about Grice's example in (2). It is in this way that these words constrain the hearer's choice of context.

What other phenomena could be analyzed as semantic constraints on relevance? In *How To Do Things With Words* Austin lists discourse connectives (like *therefore* and *moreover*) amongst a number of devices that have the function of performative verbs - that is, to indicate explicitly the illocutionary force of the utterances that contain them. Although it is easy to see the connection between *therefore* and the performative *I conclude that* , it is rather difficult to see exactly what *act* is associated with it. This is even more difficult in the case of *moreover*. But then, as Urmson (1952) suggests, perhaps many of the verbs that have been analyzed as performatives have less to do with indicating the illocutionary act that is being performed as with priming the hearer to see the 'emotional significance, logical relevance and reliability of our statements.' Thus analyzed, many so-called performative verbs have the same sort of function as the so-called cognitive verbs like those in (12), and, of course, the discourse connectives I have just been discussing.

(12)   I *think* that he has missed the train
        I *suspect* that he has missed the train
        I *know* that he has missed the train

Such verbs, says Urmson, 'do not have any descriptive sense, but rather function as signals guiding the hearer to a proper appreciation of the statement in its context'. Semantic constraints on relevance again?

Although I think Urmson was right to say that many performative verbs have this function, I believe he was wrong to say that they do not have any descriptive content. As François Recanati has shown in his book on the pragmatics of performative utterances, there is no reason to think that the utterances in (13) do not express propositions with truth values.

(13)(a) I predict that he won't come
    (b) I warn you that the path is slippery
    (c) I promise that I won't smoke
    (d) I bet $5 that she won't come

Certainly, they cannot be regarded as describing an existing state of affairs or as reporting the speaker's belief. Nevertheless they do represent states of affairs, and the speaker can be regarded as giving evidence that the state of affairs they represent obtains. What makes them special is that this evidence consists in the very act of producing the utterance. So by producing the utterance in (13a) the speaker gives a guarantee of the truth of (14a), by producing the one in (13b) he gives a guarantee of the truth of (14b), and so on.

(14)(a) The speaker is predicting that she won't come
    (b) The speaker is warning the hearer that the path is slippery

This is not to say that this is all there is to say about the interpretation of these utterances. However, it is not clear that the rest of the story is the same for all utterances that have been called explicit performatives.

Speech act theory proceeds on the assumption that the classification of speech acts plays an essential role in communication so that what is communicated by an utterance includes its assignment to a particular speech act type. So a speaker who is making a prediction must communicate the fact that he is making a prediction; a speaker who is issuing a warning must communicate the fact that he is issuing a warning, and so on. In each case communication will succeed only if the hearer identifies the type of speech act being performed. Sperber and Wilson have shown that this assumption can only be maintained in some cases. It is true that a hearer will understand an utterance intended as a bet only if she recognizes that this is indeed what the speaker is doing. That is, understanding such an utterance depends on being able to produce the description in (15). Similarly, a hearer will miss out on much of the intended relevance of an utterance meant as a promise unless she recovers the description in (16).

(15)   The speaker is betting that ......
(16)   The speaker is promising that ......

Betting and promising are what Sperber and Wilson call communicated acts.
       In contrast, consider predicting and warning. The hearer does not have to recover the description in (14a) in order to understand (17) as a prediction.

(17)   She won't come

She will understand (17) as a prediction provided that she recognizes that the speaker is committing himself to the truth of a proposition about the future. Indeed, it is in virtue of understanding the utterance in this way that she is able to provide the description in (14a). In other words, the main relevance of the performative in (13a) lies in the embedded proposition. Similarly, a speaker who intends (18) as a warning does not intend the hearer to recover the description in (14b). He simply intends the hearer to process it in a particular way. More specifically, he expects her to recover certain types of contextual implications - ones that have to do with the dangerous or unpleasant consequences of the situation the utterance describes.

(18)   The path is slippery

Once again, if the hearer does recover the description in (14b), it is as a result of having understood the utterance. The main relevance of (13b) lies in the embedded proposition.
       But if the main point of the utterance lies in the embedded proposition, what is the function of the performative verb? Why does the speaker of (13b) communicate the proposition in (14b) if the main relevance of the utterance lies in the embedded proposition? As Urmson has shown, this question arises not only for the so-called performative verbs like *warn* and *predict*, but also for the psychological verbs like *think* and *know*. The answer suggested by Sperber and Wilson is that the speaker must be understood to be performing two distinct acts of communication, one act being designed mainly to help with the processing of the other. In other words, one act of communication constrains the interpretation of another. If this

sounds strange, think of a cookery demonstrator who as she adds the flour says, 'I am now adding the flour'. The interest of utterances like (13a) and (13b) lies in the fact that the two acts of communication are simultaneous. However, in discourse we often find examples of utterances whose relevance lies entirely in the way they constrain the interpretation of the surrounding text. Consider, for example, (19) and (20).

(19)   Do you see that building over there? Well, apparently, it is on the site of a Roman theatre.

(20)   Remember the man who bought your car? I saw him at the university today and .....

In each case the point of the first segment is to ensure that certain information is accessible for the interpretation of following utterances. As in the case of the non-truth-conditional constraints on relevance, the phenomenon is a consequence of the speaker's goal in communication - achieving optimal relevance.

REFERENCES

Austin, J.L. 1962. How To Do Things With Words. Oxford: Clarendon Press.

Blakemore, D.L. 1987. Semantic Constraints on Relevance. Oxford: Blackwells.

Blass, R. 1985. Coherence, Cohesion and Relevance, unpublished paper, University College London.

Grice, H.P. 1989. Logic and Conversation, in H.P.Grice, Studies in the Way of Words. Cambridge, Mass: Harvard University Press.

Karttunen, L. 1974. Presupposition and linguistic context. Theoretical Linguistics 1, 181-194

Karttunen, L. & P. S. Peters 1975. Conventional implicature in Montague grammar. Proceedings of the First Annual Meeting of the Berkeley Linguistics Society, 266-278.

Recanati, F. 1987. Meaning and Force. Cambridge: Cambridge University Press.

Stalnaker, R. C. 1974. Pragmatic presuppositions. In Munitz, M.K. & D.K. Unger (Eds.) Semantics and Philosophy. New York: New York University Press.

Stalnaker, R.C. 1975. Indicative conditionals. Philosophia 5, 269-286.

Sperber, D. and Wilson, D. 1986. Relevance: Communication and Cognition. Oxford: Blackwells.

Urmson, J.O. 1952. Parenthetical Verbs. Reprinted in A. Flew, ed. 1966, Essays in Conceptual Analysis. London: MacMillan.

The Role of Conversational Implicature in the Early Grammaticalization of the English Perfect[1]

Kathleen Carey
University of California, San Diego

1.0 Introduction

In recent years, many linguists have demonstrated that pragmatics plays a major role in language change. Traugott and König (in press) show how semantic change can be brought about by the conventionalization of conversational implicatures through pragmatic strengthening; they show how this process explains the change from the temporal to the concessive meaning of *while* in English. Traugott (1989) shows how this same process accounts for the shift from the deontic to the epistemic meaning for the English modals *shall, will* and *must.* Horn (1984) accounts for cases of lexical narrowing and broadening as a result of his R Principle (Say no more than you must) and his Q Principle (Say as much as you can). This paper will demonstrate that the early grammaticalization of the English perfect is yet another example of the role of conversational implicature in instigating semantic change.

Before turning to the English perfect, it will be useful to cite an example that demonstrates how the conventionalization of invited inferences can result in semantic change. One of the most discussed examples in the literature is the semantic shift in which temporal connectors take on the additional semantic function of causal connectors. Traugott and König show that Old English *siþþan* 'after, since' was used as a temporal connector before it was used as a causal connector. However, during the period when it was primarily used as a temporal connector (850-1050), one can find examples such as (1), which include a causal implicature.

(1) Orosius 156.11

þa siþþan he irre wæs & gewunded, he ofslog micel þæs folces.
'Then, after/since he was angry and wounded, he slaughtered much of that troop.'

Traugott and König note that the presence of the adjective *irre* 'angry' favors a causal rather than a temporal interpretation if the listener assumes that the speaker is obeying the Gricean maxim of relevance. Over time, the causal inference becomes conventionalized as one of the possible meanings of *since.* In Modern English, *since* can be used strictly as a causal connector with no temporal reference, as is demonstrated by the following example: *Since you are not coming with me, I will have to go alone.* (Traugott and König).

2.0 Synchronic Facts about the English Perfect

The purpose of the present paper is to show that conversational implicature is also involved in the early grammaticalization of the English perfect. Before examining the perfect diachronically, let's first briefly review what some linguists have said about the Modern English perfect from a synchronic perspective. Many linguists have commented on the seemingly schizophrenic nature of the English

perfect, dividing the perfect into different types. The types listed below and their corresponding examples are a summary taken from Brinton (1988). [2]

A. Resultant State Perfect: past action with present results.
I have eaten lunch. (and am therefore not hungry now)
He has caught a cold. (and cannot come to work)

B. Perfect of Experience: a situation that occurred once or more in the past within a span of time construed by the speaker as continuing up to the present.
I have been abroad several times.
I have read that novel.

C. Continuative Perfect: a situation which began in the past and persists until the present.
We have known him since he was a child.
He has sung in the choir for years.

From a strictly synchronic perspective, one would want to know whether these types represent different meanings of the perfect or can be captured by a single definition. Bauer (1970) argues that there is a single grammatical meaning for the perfect: 'What the perfect qua perfect does in fact express is simply this: the action is viewed, not as a past event, but as being an accomplished fact at the moment of speaking, having taken place, once or repeatedly, within a span of time that is not conceived as separated from the moment of speaking.' (p. 190). Bauer goes on to explain how all of the types of the perfect are consistent with this meaning and how lexical verb semantics interacts with this meaning. His observations are relevant for this paper because lexical verb semantics turns out to be significant for the perfect diachronically as well as synchronically.

In particular, he distinguishes between two kinds of verbs: those that imply that the action reaches a goal or conclusion (telic), e.g. *buy, sell, achieve, contrive, persuade,* and those that don't imply a goal or conclusion (atelic), e.g. *love, live, walk.* Telic verbs can form Resultant State Perfects, such as *I have persuaded him,* because there is a goal or resultant state inherent in the semantics of the verb. In contrast, atelic verbs do not participate in Resultant State Perfects. When used without an adverbial indication of time, perfects with atelic verbs are typically interpreted as Perfects of Experience, as in *I have lived in Spain.* When used with a time adverbial, atelic verbs typically form Continuative Perfects, such as *I have lived in Spain for three years.* Telic verbs can also participate in Perfects of Experience, such as *I have persuaded him many times in my life,* and in Continuative Perfects when time adverbials are present, such as *He has told that joke for many years.* Note, however, that Continuative Perfects with telic verbs are construed as a series of events leading up to the present rather than as a single, continuous event.

3.0 The Diachronic Development of the English Perfect

Linguists are in general agreement that the English perfect originated in constructions with transitive verbs in which the participle was adjectival. Mitchell

(1985, p. 292) states that '...we need, I think, have little hesitation in accepting the orthodox view that in the original form of the periphrasis the participle was inflected and adjectival.' In other words, the Old English precursors of the perfect were semantically equivalent to the Modern English stative construction, as in *I have the car washed*, in which the past participle *washed* serves as an adjectival complement. Unlike Modern English, the order of the participle and the object in Old English could not be relied upon to distinguish between the adjectival and the perfect interpretation. (Throughout the rest of this paper, I will use the term *have + participle* construction to refer to any clause with an inflected form of *habban* and a past participle, regardless of the order of the constituents.) In addition, the participles in *have + participle* constructions appeared both with and without inflection: in (2) below, *geswene* 'seen' is inflected but *gehyred* 'heard' is not.

(2) ÆCHom I, 39 578.24 (Traugott, to appear)
    Fela Godes wundra we habbað gehyred and eac geswene.
    'We have heard and also seen many of God's wonders.'

Given that the perfect had its origins in an adjectival construction, the next question is exactly how does the semantic change from the adjectival to the perfect construal come about? The next part of the paper will present a hypothesis of how this semantic shift occurred based on the conventionalization of invited inferences. In presenting this hypothesis, the following questions will be addressed: (i) of the three types of the perfect mentioned above, which was the first to appear? (ii) were the first uses of the perfect restricted to a particular semantic class of verbs? (iii) how did the perfect meaning become conventionalized? Unfortunately, the earliest stages of the development of the perfect predate the Old English texts; however, the discussion will include the extent to which the Old English data supports or does not support the hypotheses in regard to later stages of development.

3.1 Contextual Constraints on the First Perfects

When semantic change arises from the conventionalization of invited inferences, some aspect of the context in which the expression was used with its old meaning becomes indexed and over time becomes part of the new meaning of the expression itself. This type of account entails that the very first uses of the new meaning, i.e. with invited inferences to the new meaning, will arise from contexts in which the old meaning was used. In the case of *siþþan* 'since', the first uses with a causal inference appeared in contexts where *siþþan* was functioning as a temporal connector.

Likewise, an invited inference account would predict that the first uses with the perfect meaning emerged from contexts in which the adjectival construction was already being used. By examining the constraints on the context imposed by the adjectival construction, one can hypothesize which of the three types of the perfect discussed earlier was the first to appear. Consider the following ungrammatical sentence:

(3)**I have my room cleaned, but you messed it up.*

(3) is ungrammatical because the adjectival construction is a statement about the present state of the object; the object must necessarily be in the state denoted by the past participle at the time of reference, in this case, the moment of speaking. The Resultant State Perfect (with transitive verbs) is the only type that must necessarily satisfy this constraint. Perfects of Experience need not and typically do not satisfy this constraint; the only relevant resultant state in a Perfect of Experience is in the subject rather than the object: in (4) below, Sheila, the subject, is construed by the speaker as being in the experiential state of having dyed her hair blonde once in her life.

(4) *Sheila has dyed her hair blonde.* (With Perfect of Experience reading)

Continuative Perfects need not and typically would not satisfy the resultant state constraint either: Continuative Perfects with atelic verbs never imply a resultant state in the object because of the inherent semantics of atelic verbs. Continuative Perfects with telic verbs, as in (5) below, cannot be ruled out; however, it is highly unlikely that the first uses of the perfect would refer to multiple rather than single resultant states.

(5) *Grandma has baked a pie every Saturday for the last 20 years.*
    (Continuative)

From this analysis, we can hypothesize that the first perfects should be of the type Resultant State, and consequently with telic verbs. Atelic verbs are clearly incompatible with the adjectival construction, as is indicated by the ungrammaticality of *I have the Italian shoes wanted.* This hypothesis is supported by data from Old English prose; an analysis of a total of 167 *have + participle* examples from Alfred (c. 850) and Ælfric (c. 1050) reveals that 165 were with telic verbs.[3]

3.2 The Shift from the Adjectival to the First Perfect Construal

The next issue to be addressed is how the adjectival constructions developed into Resultant State Perfects. Before addressing this question, one must first understand how the two constructions are semantically different.

The primary semantic difference between the adjectival construction and the first uses of the Resultant State Perfect has to do with the syntagmatic relations of the elements. In the adjectival construction, *have* acts as a main verb, designating a relation between the object and the subject. The focus of attention in the adjectival construction is not on the relation of possession/control but rather on the state of the object. The past participle acts adjectivally by designating the present state of the object, which is the final state of a past process. In the first uses of the Resultant State Perfect, *have* designates a relation between the subject and the past process performed on the object. The past participle is verb-like in that it designates a past process.[4] Importantly, these differences between the adjectival construction and the first Resultant State Perfect do not correspond to situations with different truth conditions but rather to different construals of the same situation.

A truth-conditional semantic difference between the adjectival and the perfect construction is that, in the perfect construction, the subject must necessarily be the agent of the process. Here, I am using the term agent in the broad sense to refer to consciously controlling participants. Example (6) demonstrates that the subject of an adjectival sentence need not be the agent of the process:

(6) A: *John has Meow bathed and ready for the cat show.*
    B: *John bathed Meow?*
    A: *No. Mary did. Meow would never let him do that.*

However, agency of the subject is part of the meaning of the modern perfect construction, and not just an implicature: note that in example (7) the agency of the subject cannot be overruled without sounding contradictory.

(7) A: *John has bathed Meow; she's ready for the cat show.*
    B: *John bathed Meow?*
    A: * *No. Mary did.*

In most contexts in which the adjectival construction would be uttered, the subject is the agent of the process, and consequently, these contexts satisfy the truth-conditional requirements for the perfect meaning. In a particular utterance, the additional necessary shift from the adjectival to the perfect construal could be brought about by implicature. In example (8) below, it is possible to interpret the construction as purely adjectival or as including an implicature to the perfect construal, as I have indicated with the two possible glosses.

(8) ÆCHom I, 31 458.18 (Traugott, to appear)
    Ic hæbbe gebunden þone feond þe hi drehte.
    'I have [bound that enemy/that enemy bound] that afflicted them.'

However, the possibility of an implicature does not necessarily indicate that the implicature will be conventionalized as a new meaning for the construction. In the case of the *have + participle* construction, whenever the object is tangible and changed by the verbal process, i.e. an external object, it bears the resultant state designated by the participle; consequently, the purely adjectival meaning is always possible. Since the adjectival meaning is also the original meaning, it is difficult to see how examples such as (8) would consistently lead to an inference with the perfect construal and consequently bring about the conventionalization of the perfect meaning.

3.3 The Role of Mental State Verbs in the Conventionalization Process

However, the situation is much clearer with mental state verbs. Benveniste (1968) claims that verbs such as *understand, discover, realize, notice,* and *see* were the first to favor the rise of the perfect. He argues that sensory-intellect verbs such as these are specially suited for this task for the following reasons: (i) the subject must necessarily be the agent of the process, and (ii) although the participle still

grammatically modifies the object, the change in state brought about by the process involves the subject rather than the object.

Presuming the achievement rather than the stative use of a mental state verb, if an object stands in a *heard, understood, discovered,* or *seen* relation with a subject, the subject must be the agent of the process.[5] Consequently, the *have + participle* constructions with mental state verbs entail the truth conditional requirements for the perfect. The development of the constraint that the subject must necessarily be the agent of the process is crucial for the development of the perfect; it permits the syntagmatic realignment in which the verb consists not just of *have* but of the *have + participle* complex.

Benveniste's second point (ii) explains how the uses with mental state verbs bring about the shift from the adjectival to the perfect construal. Consider the following Modern English sentence: *Now that I have that paper understood/figured out, I can start on the next one.* In this example, the participle grammatically functions as an adjectival complement, as is indicated by the stative syntax of the first clause. However, in a context in which this statement would be uttered, the listener would typically infer that what is relevant is not the *understood* state of the paper but rather the change of state of the subject, i.e. that the subject has completed the process of understanding. In the inference-based construal, the participle functions more like a verb by designating a past process. In this way, the particular semantics of mental state verbs leads to an inference that results in a shift from the adjectival to the perfect construal.

In several examples from the Alfred period of Old English (c. 850), the object of the mental state verb was a proposition rather than a tangible object such as a paper. When the object is a proposition, as in (9) below, and therefore an object only in the most abstract sense, the implicature to the perfect construal is even more apparent.

(9) Bo 36.107.32
    ic hæbbe nu ongiten þæt þu eart gearo to ongitanne mina lara;
    'I have now understood that you are ready to understand my teachings;'

Verbs of reporting such as *tell, explain,* and *say* also play a role in the conventionalization of the perfect meaning. As with mental state verbs, *have + participle* constructions with verbs of reporting typically include an implicature to the perfect construal. Consider the following example from Old English:

(10) Or 1 1.26.26
    Nu hæbbe we ymb Affrica Landgemæro gesæd.
    'Now we have talked about the African land.'

In (10), the listener infers that the speaker does not want him/her to focus attention on the *talked about* state the African land; to do so would be to violate the Gricean maxim of relevance. The listener infers instead that what is relevant is the state of the discourse situation, in particular, the fact that the subject has completed the process of talking about Africa. In this way, verbs of reporting are similar to

mental state verbs by typically including an implicature in which the participle has a verbal rather than an adjectival function.

3.4 Evidence from Old English

In order to test the hypothesis regarding the conventionalization process, 167 *have* + *participle* examples from Old English (Venezky and DiPaolo, 1980) were categorized by the semantics of the verb participle. 84 of the examples were taken from the prose of Alfred (c. 850) and 83 of the examples were taken from the prose of Ælfric (c. 1050). The data includes both past and present tense forms of *habban*; the majority of the examples were first and third person.

An analysis of these *have* + *participle* examples indicates that 54 (32%) of the examples were with mental state verbs, and 32 (19%) were with verbs of reporting; together, they comprise over 50% of the data. Of the remaining examples, 66 (40%) were with verbs that designated a change of state in the object, e.g. *belocen* 'locked', *gedon* 'done', and *gebunden* 'bound'. It is difficult to determine whether any particular example from this category had the perfect or the adjectival construal; my hypothesis is that the majority of them had the adjectival construal, particularly the examples from the earlier Alfred period.[6]

The remaining 15 examples (9% of the data) included verbs whose participles do not designate a change in state of the object, or at least a perceptually verifiable one, verbs such as *gebroht* 'brought' and *betæht* 'taught'. Examples in the last category seem to be instances in which the perfect construal is extending to new verbs, that is verbs that are neither mental state nor reporting. Interestingly, 12 of the 15 examples in this category are from the later Ælfric period; in fact, some of these verbs, such as *gebroht* 'brought', do not appear anywhere in my data sample from the earlier Alfred period. This data suggests that once the mental state and reporting verbs have conventionalized the perfect meaning as a legitimate meaning of the *have* + *participle* construction, it can start to extend to other verbs.[7]

In order to validate the above hypothesis, one would need data from later periods to conclusively demonstrate the extension to other semantic classes of verbs. This paper represents the initial results of a work in progress, and I have not yet analyzed data from later periods. However, it is clear that by Middle English, the perfect had extended not only to atelic, but also to stative verbs, as example (11) below demonstrates. Also, by Middle English, examples of the Perfect of Experience, such as (12), had emerged:

(11) Curs. M 7994 (Visser 1973, p. 755)
His hert has ever ben þe with.

(12) Chaucer C.T. A3079 (Visser, p. 755)
Your owne knight, That serveth you with wille, herte and might, And ever hath doon.

4.0 Conclusion

Applying pragmatic principles to semantic change has led to the following hypotheses regarding the development of the English perfect: (i) The first perfects

should be Resultant State Perfects with telic verb participles. (ii) Verbs with external objects may involve an implicature to the perfect construal, but they will not play a significant role in conventionalizing the new meaning. (iii) Mental state verbs conventionalize the new meaning by entailing that the subject is the agent of the process. Both mental state verbs and verbs of reporting conventionalize the new meaning by typically including an implicature to the perfect rather than the adjectival construal. (iv) After the conventionalization process has occurred, the perfect meaning can be extended to other semantic classes of verbs, including verbs with external objects.

If the hypotheses presented in this paper are accurate, they indicate that the early grammaticalization of the English perfect is yet another example of the role of conversational implicature in semantic change. The hypothesis regarding the role of mental state verbs in conventionalizing the implicature suggests the following general principle: An implicature associated with a grammatical construction is more likely to be conventionalized if a subset of the uses of that construction entails some part of the new meaning; recall that in the case of the perfect, the uses with mental state verbs entailed that the subject was the agent of the process.

Notes

[1]I would like to thank the following people for their comments on earlier versions of this work: Claudia Brugman, Gilles Fauconnier, Mary Hare, Rob Kluender, Yuki Kuroda, Ron Langacker, Ricardo Maldonado, Linda Manney, Teenie Matlock, Errapel Mejiás-Bikandi, Guenter Radden, Jo Rubba, and Maura Velazquez. Special thanks go to Suzanne Kemmer for comments on several versions of this work. The usual disclaimers apply.

[2]Brinton also includes a fourth type, the Perfect of Recent Past, which can be considered a subtype of the Resultant State Perfect. I have omitted it here.

[3]The two *have + participle* constructions with atelic verbs are nonstative and include a time adverbial. The specification of the time adverbial adds an implied goal not present in the verb itself, so that the entire construction can be construed as telic although the verb itself is atelic.

[4]It is still unclear to me exactly how the notion of relevance applies to the first uses with the perfect construal: in particular, whether the past event must be relevant to the subject or instead to some aspect of the speech situation or its participants. See Langacker (1990) for a discussion of related issues.

[5]Someone other than the subject can be the agent of a mental process only if explicitly specified in a causal construction such as *The king has the book understood by his scribe.*

[6]Errapel Mejiás-Bikandi has provided synchronic evidence from Modern Castilian Spanish that supports the hypothesis regarding the role of mental state and reporting verbs. In addition to the fully grammaticalized *haber* perfect, Castilian Spanish has what appears to be a partially grammaticalized perfect with *tener* as the auxiliary. Consider the following examples:

Tengo entendido que Vd. sabe la respuesta.
'I have it understood that you know the answer.'

Te tengo dicho que no uses el coche.
'I have it told to you not to use the car.'

Native speakers feel that these sentences with *tener* focus more attention on the state of the object than their *haber* counterparts. This difference is to be expected, since the perfect meaning has not yet been conventionalized with the *tener* perfect. The *tener* perfect is restricted to mental state verbs, verbs of reporting and verbs that denote final states, such as *preparado* 'ready' and *hecho* 'done'. Transitive verbs without objects and intransitive verbs cannot appear in this construction. Also, the participle obligatorily agrees with the object in gender and number (and in some cases, can appear in front of the object):

Tengo   la película   vista.
I have   the movie   seen-F
'I have [seen the movie/the movie seen].'

[7]Three of the examples were with intransitive verbs; in two of these, the verb was *faran* 'go', and in the other example, the verb was *restan* 'rest'. In understanding this somewhat puzzling fact, it is worthwhile to note that a *be* perfect was developing with mutative intransitive verbs such as *faran* and *cuman* 'come' throughout the same period that the *have* perfect was developing with transitive verbs. Perhaps, the influence of the *be* perfect partially explains the early appearance of the mutative intransitives in *have* + *participle* constructions. However, Mitchell (1985, p. 290) notes that this does not explain the early appearance of activity verbs such as *restan* and *settan*. I have no coherent explanation for these facts; however, it is significant to note that all of the early examples with activity verbs included a time adverbial. As was explained in endnote 3, the addition of a time adverbial provides an implied goal, so that the event is telic although the verb is inherently atelic.

References
Bauer, Gero. 1970. The English Perfect Reconsidered. Journal of Linguistics, 6:2:189-198.

Benveniste, Emile. 1968. Mutations of Linguistic Categories. Directions for Historical Linguistics, ed. by W.P. Lehmann and Yakov Malkiel, 85-94. Austin: University of Texas Press.

Brinton, Laurel. 1988. The Development of English Aspectual Systems. Cambridge: Cambridge University Press.

Horn, Laurence. 1984. Toward a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature. Meaning, Form and Use in Context, ed. by Deborah Schiffrin, 11-42. Washington, DC: Georgetown UP.

Langacker, Ronald. 1990. Subjectification. Cognitive Linguistics, 1:5-37.

Mitchell, Bruce. 1985. Old English Syntax. Vol. I. New York: Oxford University Press.

Traugott, Elizabeth. 1989. On the Rise of Epistemic Meanings in English: An Example of Subjectification in Semantic Change. Language 65:31-55.

Traugott, Elizabeth. To appear. Old English Syntax. The Cambridge History of English, Vol I, ed. by Richard Hogg. Cambridge: Cambridge University Press.

Traugott, Elizabeth, and Ekkehard König. In press; will appear 1990. The Semantics-Pragmatics of Grammaticalization Revisited. Approaches to Grammaticalization, ed. by Elizabeth Traugott and Bernd Heine. Amsterdam: Benjamins.

Venezky, Richard L., and Antoinette DiPaolo Healey. 1980. A Microfiche Concordance of Old English. Dictionary of Old English Project, Center for Medieval Studies: University of Toronto.

Visser, T. Th. 1973. An Historical Syntax of the English Language. Part I, Vol. 2. Leiden, The Netherlands: E.J. Brill.

Gricean Maxims and Reading Instruction
Charles Elster
Purdue University

Introduction. This paper is an application of linguistic pragmatics to the area of literacy education. My goal is to explore how the Gricean concept of conversational cooperation and the maxims which support it (Grice 1975) apply, or apply in unusual ways, or fail to apply to the reading tasks which beginning readers encounter in classrooms.

My thesis will be that while the cooperative principle and maxims do govern classroom reading activities, as they govern other communicative events, the maxims operate under constraints that apply only in the classroom. These unique constraints have less to do with differences between spoken and written language than with differences between classroom discourse, whether spoken or written, and discourse outside of classrooms.

Theoretical Background. Educational research in the past twenty years has described how classroom discourse differs from discourse outside of classrooms (Sinclair & Coultard 1975, Mehan 1979, Gumperz 1981, Heath 1982). This research shows for example that in classrooms topics and turn taking are often centralized and teacher-controlled, known answer questions are common, and answers to questions are often evaluated by teachers.

Other research has focused on the pragmatic demands of the written texts and of texts used in reading instruction (Griffin 1977, Morgan & Green 1980, Fillmore & Kay 1983, Lakoff no date).

It seems clear that Grice's cooperative principle and maxims apply in the literate transaction between readers and writers (Tierney & LaZansky 1972). That is, readers read with the assumption that writers intend to communicate, that they wish to be efficient in their communication, and that a relationship of relevance exists between the sentences in an extended written discourse. Consequently, readers interpret items in written texts with reference to the assumed intentions communicated in the text. This interaction between reader and writer through the text relates to what Fillmore (1974) has called the INTERNAL CONTEXTUALIZATION of discourse, 'the possible properties of the world depicted or implied by the text ... the worlds in the imagination of the creator and interpreters of the text' (V-5). Readers are justified in making inferences about this text-internal world based on the assumption that conversational maxims are in effect.

Literature does have unique demands, but it does not violate the cooperative principle. The obscurity of certain poetry reflects unique communicative intentions. We do not judge works of fiction by the same standard of truthfulness (QUALITY) as conversations, but readers do apply tests of plausibility and internal consistency to works of fiction.

But in addition to the creation of an imaginary world, readers and writers use literacy as part of their participation in society: publishing and reading newspapers to keep people informed, technical manuals to guide the actions of workers, or paperback novels to keep authors, publishers and printers employed and readers entertained during their leisure time. These social relations relate to the EXTERNAL CONTEXTUALIZATION of written discourse, 'the worlds in which the text can be appropriately used'.

It is the external contextualization of texts which will be the focus of my discussion. In classrooms the external context of literacy is very much in evidence, since there texts are not read by individuals in seclusion for their own pleasure, but by groups

of children and adults to fulfill a variety of classroom tasks. I would like to explore how Gricean maxims apply to reading tasks in classrooms. I will be showing what children need to learn in order to use texts appropriately in school.

Data and Method. The data presented here comes primarily from a year-long observational study of reading-related activities of first graders in two Bay Area public school classrooms. The classroom data includes transcripts of audiotape recordings of children working with teachers in reading groups and on their own doing workbook exercises. Additional data comes from widely-used basal readers and workbook pages completed by children in a midwestern first grade classroom.

Variety of Classroom Tasks. Two types of texts will be of primary interest: the stories found in basal readers, and the pages of exercises found in basal workbooks. My contention is that elementary school children learn to interpret written texts in terms of the goals associated with typical classroom tasks. A typical elementary classroom day consists of a succession of different social-instructional contexts, each with a different set of texts, different tasks and norms of talk and reading, and sometimes different participants. Some of these typical classroom contexts are Rug Time, when class business is transacted, Reading Time, when children work in teacher-led groups and independently on tasks associated with basal workbooks, and Sustained Silent Reading Time, when children read independently in books of their own choosing.

From a Gricean perspective, children and teachers adjust their goals for different tasks, and these adjustments affect the type of cooperation required and the interpretation of conversational maxims. That is, a response or interpretation that might be appropriate in one setting (relevant or quantitatively sufficient) might be viewed as inappropriate and uncooperative in another. Conversational cooperation is dependent on a shared understanding of the goals implicit in each context.

I will now present illustrations of how three Gricean maxims - QUANTITY, RELATION and MANNER - apply to the contextualization of written texts in elementary school classrooms.

Quantity. The maxim of QUANTITY states that one should be as informative as necessary in a given situation. I will examine primer texts and workbook texts as they relate to QUANTITY. From the standpoint of the internal contextualization of texts, primer texts taken on their own might be considered uncooperative due to their artificially controlled language and their dependence on accompanying pictures. To be quantitatively cooperative, primer texts need to be interpreted in connection with their accompanying pictures and the talk and tasks in reading lessons.

Example 1: The Fire Dog (Clymer & Venezky 1982)

> '"Have you come to see us?" the man asked.
> "We have", she said.
> "We have come to see what you do here."'

Taken on its own, these opening lines of the story "The Fire Dog" do not contain sufficient information to interpret the pronouns <u>you</u>, <u>us</u>, <u>we</u>, and <u>she</u>, the definite expression <u>the man</u>, and the deictic expressions <u>come</u> and <u>here</u>. It is only when the accompanying picture is added, showing a women a four children talking to a man in front of a fire station, that the text becomes somewhat clearer. The text is clarified further by the teacher-led discussion, which is scripted in the teacher's manual with directions

such as 'As pupils study the picture..., explain that the storybook children and their classmates and their teacher are on a field trip to a neighborhood fire station. ... Aid pupils in recognizing the people they know... Ask: What does the man mean by <u>us</u> when he asks, "Have you come to see us?"'.

Workbook exercises also demand a special interpretation of the maxim of QUANTITY. In Example 2 (Alvermann et al. 1989), a first grader completed a worksheet exercise by using all the possible responses from the top of the page to fill in the sentence blanks. That is, he used his knowledge that each word at the top of the page should be used in one sentence. However, after deciding that the word <u>mix</u> completed the second sentence, <u>'I will get paint mix', said Carlos.</u>, he apparently filled in the wrong answer to the last item by writing in the remaining word <u>today</u> without reading the resulting sentence: <u>'You can today blue paint and yellow paint,' said Max.</u> (The original responses are visible in dotted lines underneath the solid line corrections.)

In Example 3 (Alvermann et al. 1989), the same child answered the questions correctly and with sufficient information for normal conversational exchanges - that is incorporating new information, or comments, and leaving out given information, or topics. But the child's responses were judged incorrect by the classroom standards of QUANTITY, which required the use of "complete sentences." Under the teacher's direction, the child corrected items 1 and 2. Interestingly, item 2 was made into an acceptable sentence by removing the conjunction <u>because</u>, the logical tie between response and question. These worksheet examples demonstrate the incompatibility of judgments of QUANTITY in the classroom and outside it.

Relation. The maxim of RELATION states that one should be relevant to the communicative task at hand. Within written texts, an assumption of coherence ties together sentences. For example, temporal and causal ties are assumed to account for the relevance of successive sentences.

In terms of their appropriate use, classroom texts must be interpreted in relation to the task at hand in any given context, and these external tasks will determine the relevance of texts and their parts. For example, in order to successfully interpret and use worksheets, children must not only read the worksheets literally, but interpret what they read and see in terms of the typical tasks associated with them. Example 4 occurred while two first graders were working side by side on a worksheet.

Example 4. Independent Work

Synopsis: Alex and Meera are doing a cut and paste work sheet, where pictures of objects must be pasted in <u>sh</u> or <u>ch</u> areas depending on their initial digraph sound. Well into the activity this exchange occurs:

Alex:	I can't find another one that goes with this.
Meera:	... Ship.
Alex:	... I thought it was boat.
Meera:	But. See. To make it go, it had to be ship. Because <u>SH</u>.

In Example 4, Meera interpreted the picture as a representation of the word <u>ship</u> because she realized its relevance to the workbook task of distinguishing two digraph sounds. Alex's interpretation of the picture as representing <u>boat</u> did not take into account the relevant contrast. This example shows how children learn to interpret workbook pictures in terms of typical tasks which test discrimination of letter sounds.

During reading lessons, pages of text are often read to achieve the goal of answering teachers' questions. Children perform this reading task most successfully when they have in mind the typical functions of reading within pre-formatted lessons. Example 5 shows how a teacher's request to label a picture actually serves, within the lesson format, to introduce the title of the selection to be read.

Example 5. Reading Group

Synopsis: This exchange takes place as children are preparing to read a one page poem entitled "Clay" (Aaron et al. 1983). The picture shows a girl making toys out of clay.

Teacher:	What's this girl look like she's working with.
	(points to picture)
Talya:	Toys.
Teacher:	No, she's making them. She's making her own toys.
Talya:	Clay.
Teacher:	She's using clay. And that's the title of the poem.

In this excerpt, Talya's first response to the teacher's question is based on her examination of the pictures. Although it is literally correct, the teacher evaluates this response as inappropriate, since it is not relevant to the unstated goal of deducing the poem's title and overall topic. Talya's second response represents a second attempt to be relevant to the topic at hand. The teacher evaluates this response as appropriate.

A final example of how the maxim of RELATION operates during reading lessons involves adult responses to reading errors (miscues) and hesitations during oral reading by children. If we assume that the shared task during reading lessons is to read through a given selection, we might expect a competent adult reader to assist a child by reading words the child is having difficulty with. However, if we view the task as one of assisting children in becoming independent readers, then the appropriate response is to assist the children in marshalling strategies for decoding the word themselves. Adult responses to oral reading errors during a cooking activity are presented in Example 6.

Example 6. Cooking Center

Synopsis: Lydia (a parent) is copying a baking recipe from a
book onto a display card as she directs several children in mixing.

Lydia:	Here's a tricky word. (copying "KNEAD")
Sandra:	"Ke:n:d. Ke:n:d." (reading)
Lydia:	The K is silent.
Sandra:	"Nend. Nend. Ed. Nended. Nend. Nend."
Lydia:	Instead of trying to figure it out by what it looks like, think about what we're going to do to it now.
Cathy:	Knead it.
Sandra:	"Knead".
Lydia:	"Knead". Yea.

In this episode, the refusal of the adult (Lydia) to pronounce the word knead for the children can be interpreted as an uncooperative move in the task of decoding the recipe together. But this refusal and providing graphophonemic and semantic clues to decoding are cooperative moves in the task of training children in independent word decoding.

Manner. The maxim of MANNER states that one should be clear and orderly relevant to the task at hand, avoiding obscurity and ambiguity. There might be variations in what counts as clear and orderly depending on different tasks. For example, some modern poetry may be interpreted as too obscure and ambiguous, unless one assumes that ambiguity contributes to communicative intentions which differ from those of ordinary conversation.

In terms of classroom reading tasks, clarity and orderliness must be judged in terms of localized goals. As I showed in the discussion of QUANTITY, primer texts are unique in the way readers must construct an imaginary world. They are also unique in how they are used. While they are often in the form of stories, their function in the classroom is not to provide pleasure during leisure time, but to provide practice in reading for beginners and a basis of evaluation of children's reading for teachers. Thus, when children read primer texts during reading lessons, their conversational contribution, in the form of how they read and how much they read, must be appropriate to the instructional goals of reading groups and not to the pleasure-seeking goals of reading literature.

In reading groups, children learn appropriate manners of reading: to read orally or silently, in chorus or individually, and they also 'follow along' while the teacher or another child reads. For example, during the initial read-through of a basal story, a teacher instructs the children in the appropriate manner of reading:

Example 7. Reading Group

Teacher: Read the first two sentences to yourself.(points to text) Put your thumb
                up when you're finished, hand up if you need help.

The following example, also from a round-robin reading activity, shows the consequence of a child's inappropriate manner of reading:

Example 8. Reading Group

Synopsis: During round robin reading the teacher discourages Talya from paging ahead in her book.

Talya: (sings, flips ahead in book)
Teacher: Talya, QUIT flipping your book. When we're in a GROUP you have to
                read differently than when you're reading by yourself. You can't just turn
                pages right off cause sometimes there's other things that are important
                too. Okay? About reading?
Talya: (sighs)

In this example, Talya has difficulty conforming to the pacing of reading determined by the teacher-directed reading lesson. Her impulse to read ahead conflicted with the teachers instructional agenda. The teacher tries to explain to Talya why her behavior is inappropriate in this situation. In other classrooms contexts, such as student-directed Free Reading times, children are free to read silently and at their own pace.

Conclusions. During classroom reading activities, then, children learn to understand cooperation in terms of conversational norms unique to classrooms, and indeed unique to localized task contexts. In particular, children must interpret the maxims of RELATION

(be relevant), of QUANTITY (be as informative as necessary), and of MANNER (be clear and orderly) in terms of unusual and localized classroom tasks and goals. This reformulation of conversational maxims is a problem facing children as they seek to succeed in learning to read and write in American classrooms.

The examples presented here raise a questions: How do participants and observers of discourse, whether in or out of classrooms, recognize common purposes? What may occur in some classroom situations is a failure of teachers and children to establish and maintain shared purposes. That is, teachers may view children as uncooperative, or children may view teachers in a like manner, if they fail to construct shared cooperative frames of reference for literacy activities. Without this shared frame of reference, children's or teachers' contributions may be judged disorderly, insufficient or irrelevant to the task at hand.

Two pragmatic problems face children as they encounter classroom reading instruction. One is that reading lessons are formats which are often pre-planned by teachers to fulfill certain instructional goals, rather than the unplanned and jointly constructed conversations children are used to. A second problem facing children is that of recognizing the goal of reading activities to be instructional, rather than recreational or functional. Children may judge the texts and reading tasks they encounter in class as uncooperative if they evaluate them in aesthetic or functional terms, rather than in terms of providing instruction and practice in reading and writing.

A recent movement in literacy education, the whole language movement (Goodman 1986), builds on the assumption that literacy acquisition is quite similar to language acquisition, and that it should proceed through literacy experiences which have functional and aesthetic value rather than purely instructional ones.

From a pragmatic perspective, the whole language movement seems to offer children opportunities to engage in child-directed and open-ended interactions with natural language written texts in place of pre-planned and teacher-directed tasks associated with basal stories and workbooks. But whether whole language instruction is the answer to problems facing children in classroom reading instruction or not, it is important that teachers and researchers continue to ask themselves whether classroom reading tasks are helping children to transfer reading and other language use strategies from classroom texts and tasks to those that will engage them outside of school.

REFERENCES

Aaron, Ira, Dauris Jackson, Carole Riggs, Richard Smith, & Robert Tierney. 1983. Scott Foresman reading program. Glenview CA: Scott Foresman & Company.

Alvermann, Donna, Connie Bridge, Barbara Schmidt, Lyndon Searfoss, & Peter Winograd. 1989. Heath reading. Lexington MA: D. C. Heath & Company.

Clymer, Theodore, & Richard Venezky. 1982. Ginn reading program. Lexington MA: Ginn & Company.

Fillmore, Charles. 1974. Pragmatics and the description of discourse. Studies in syntax and semantics, ed. by Charles Fillmore, George Lakoff & Robin Lakoff. Berkeley CA: University of California.

Fillmore, Charles, & Paul Kay. 1983. Text semantic analysis of reading comprehension tests. Final NIE Report. Berkeley CA: University of California.

Goodman, Ken. 1986. <u>What's whole in whole language?</u> Portsmouth NH: Heinemann Educational Books.

Grice, H. Paul. 1975. Logic and conversation. <u>Syntax and semantics</u>, vol. 3, ed. by Peter Cole & Jerry Morgan. New York: Academic Press.

Griffin, Peg. 1977. Pragmatics and reading: Symbiosis. <u>Linguistic theory: What can it say about reading?</u>, ed. by Roger Shuy. Newark DE: International Reading Association.

Gumperz, John. 1981. Conversational inference and classroom learning. <u>Ethnography and language in educational settings</u>, ed. by Judith Green & Cynthia Wallat. Norwood NJ: Ablex.

Heath, Shirley Brice. 1982. Questioning at home and at school. In <u>Doing the ethnography of schooling</u>, ed. by George Spindler. New York: Holt Rinehart & Winston.

Lakoff, Robin. No date. <u>Expository writing and the oral dyad as points on a continuum: Writing anxiety as a result of mistranslation.</u> Unpublished manuscript. Berkeley CA: University of California.

Mehan, Hugh. 1979. <u>Learning lessons.</u> Cambridge MA: Harvard University Press.

Morgan, Jerry, & Georgia Green. 1980. Pragmatics and reading comprehension. <u>Theoretical issues in reading comprehension</u>, ed. by Rand Spiro, Bertram Bruce, & William Brewer. Hillsdale NJ: Erlbaum.

Sinclair, J., & M. Coultard. 1975. <u>Toward an analysis of discourse: The English spoken by teachers and pupils.</u> Oxford University Press.

Tierney, Robert J., & Jill LaZansky. 1980. The rights and responsibilities of readers and writers: A contractual agreement. <u>Language Arts</u>, 57, 606-613.

Example 2. Workbook page

Name _____

| Write |

green   mix   paint   stop   today

Carlos wanted to **paint** the house.

"I will get paint **today**," said Carlos.

"Can we **stop** for pizza?" said Max.

"I just want **green** paint," said Carlos.

"You can **mix** blue paint and yellow paint," said Max.

"Good!" said Carlos. "Then we will get some pizza."

Example 3. Workbook page

Pete

There he goes!

Pete can fly up in the sky.
Pete has a big, orange beak.

1. Is Pete a little boy?

☺ No, he is not.

2. How do you know that?

☺ Bees he flies

3. Where does Pete fly?

no sentence In the sky

4. What color is his beak?

no sentence Orange

5. Can you fly?

no sentence No

Invisible Meaning

Gilles Fauconnier

University of California, San Diego

The following was reported by Newsweek (Jan. 15, 1990):

Mario Cuomo: Refuses all clemency requests for first time.
De-Hortonizes himself for 1992.

The neologism found here, *de-Hortonize*, may well on that occasion, and with the exception of the present article, have made its first and last appearance in the English language. It exhibits consistent morphology, with the affixes *ize* and *de*, and the tense/person marker *-s*. Furthermore, as far as I can tell, for people reasonably familiar with the American political context, it does not pose severe comprehension problems; the understanding, in fact, is felt to be pretty much 'automatic'. Since the word is not part of the reader's lexicon when first encountered, one may ask what is involved in constructing an 'instant' interpretation, and whether the process at work is regular or linguistically pathological.

Informally, something like the following seems to be going on: *Horton* in the Newsweek context calls up a presidential campaign frame exemplified by the 1988 campaign with 'roles' for the presidential contenders (Democrat Dukakis and Republican Bush) and the 'criminal' treated with leniency (Horton). This frame is structured by various kinds of background knowledge; among other things, one of the roles, 'the contender showing leniency', is vulnerable within this frame, and in fact is defeated by the other. In the excerpt, *1992* takes us to a corresponding possible instantiation of the same frame with different values (slot-fillers) for the roles. Since the inferential structure of the frame is preserved, we get the correct inference that if Cuomo were to grant clemency requests, he would be vulnerable, and perhaps defeated, in the 1992 campaign. To *de-Hortonize* oneself, then, is to avoid fitting the relevant role in the frame; other inferences are latent: for instance, combining the explicit information that this is the first time Cuomo has not granted clemency, and the presuppositional properties of the prefix *de-*, the reader is liable to assume that until then, Cuomo did meet the conditions for the 'Horton' frame.[1]

A formally precise and theoretically anchored account of this phenomenon would clearly demand much work. For present purposes, it is sufficient to note that very rich structure and transfer functions are mobilized in the understanding of such an example on the basis of very little linguistic information (the string *prefix-proper name-suffix*). The meaning of the expression, if we choose to call it that, is for the most part invisible.

It has been argued persuasively, e.g. in Travis (1981), Nunberg (1978), Kay and Zimmer (1976), Ryder (1989), that such meaning construction, far from being exceptional, is in fact the rule for unobtrusive everyday grammatical combinations such as 'Adjective Noun', 'Article Noun', 'Noun Noun', etc. One way to put this is that much of meaning is invisible - not encoded in a systematically retrievable fashion by the linguistic

forms. And it's worth noting that while this feature is supported by an abundance of data, it runs counter to our folk theories of language and to most formal theories as well: core meaning is commonly viewed to be entirely supported by words, with special pragmatic provisions for contextual effects.

Grice's work on implicature came into linguistics in the late sixties and had a major impact on the evolution of the field; one reason was that it opened up and legitimized the study of invisible meaning, i.e. inferences that are essential for a proper understanding of what is said, and yet bear little or no connection to any manifest linguistic structure in the relevant speech situation. At the time when this occurred, the theoretical emphasis in linguistics was largely on syntax, which had been elevated to quasi-scientific status thanks to work by Harris, Chomsky, and their students; little attention was devoted to semantics for its own sake: meaning tended to come into the picture via the underlying structures of syntax, and pragmatics, if mentioned at all, was at best a convenience for dismissing embarrassing discrepancies between theoretical expectations and observed behavior. Grice unwittingly opened Pandora's box. The invisible meaning (conversational implicatures, indirect speech acts and the like) needed to make sense of language expressions was not only hard to account for; it turned out to affect syntactic distributions in important ways (cf. Ross (1975), Green (1975), Schmerling (1971), Horn (1972), Fauconnier (1975), Cornulier (1984, 1985), Anscombre and Ducrot (1983)).

But Grice and his epigones were operating within a 'classical' scheme: language expressions were first endowed with a full truth-conditional literal meaning in virtue of their structure, and only later was a derived (extended or modified) meaning produced under pressure from pragmatic communicative principles. In fact, a major appeal of the Gricean approach was that it simplified , or regularized, the literal meaning component: the discrepancies between standard logic and natural language were to be accounted for by combining a well-behaved literal meaning component with appropriate pragmatic principles.

In Gricean terms, the problem of invisible meaning takes on the following form: given that expressions come with literal meaning, *why* does this core meaning get transformed or expanded in actual communication, and *where* does the extra meaning come from? Typically, powerful Gricean notions such as the Cooperativeness Principle and the Maxims address the *why* part of this question more than they do the *where* part; apparent or real lack of relevance or of quality triggers the search for implicature but does not in and of itself determine the implicature content.

Matters appear in a different light if we take a view of meaning construction like the one mentioned at the outset in connection with the 'Horton' example. Under such a view, it is in the very nature of linguistic form to considerably underspecify meaning construction; the search for 'invisible' meaning is on from the start: context and prior discourse configurations must be invoked directly before any meaning at all, literal or derived, can emerge.[2] From that perspective, Gricean phenomena are no longer sharply different from other meaning constructions, and the role played by communicative principles (like the maxims) deserves to be reexamined.

Using data from hypothetical and counterfactual phenomena, I will suggest in the present paper that the construction of interpretations includes the following kinds of steps:
- **constraint satisfaction**: as discourse construction unfolds, each language expression that comes along can be viewed as placing strong constraints on the next step of the construction;
- **construction strategies**: to choose between the constructions compatible with the constraints (imposed by the language form), strategies operate that are based on previous

stages of the discourse configuration, on default, simplicity, and optimization principles, and on 'look-ahead' techniques;

- **pragmatic inference**: although the configurations obtained in the first two steps are themselves sensitive to pragmatic factors, they require properly pragmatic operations to be fully interpreted; such operations are apparently often based on simplified cultural and psychological models of various sorts.

1. Transfer functions in counterfactuals

I assume the familiar approach in which hypotheticals (including counterfactuals) set up new discourse domains ('spaces') which are linked to previous discourse domains by space to space mappings called 'connectors'.[3] Interesting grammatical and semantic properties follow from the fact that spaces can be linked by more than one connector. Consider the following counterfactuals:

(1) If Henry was Gorbachev, he would bomb Lithuania.

(2) If Bush was Gorbachev, he would cut the Pentagon's budget.

(3) If Henry was Lucy, he would have blue eyes.

There is a wide variety of interpretations that such sentences could take in context. For present purposes, I will single out the following:

(1): a statement about Henry, e.g. his aggressivity, or his opinion about what Gorbachev should do;

(2): what Gorbachev might do if he were running the U.S.;

(3): what Henry would look like if he happened to be Lucy; *(3)* might be uttered by someone who has just met Henry and suspects that he might be Lucy in disguise.

In all three examples, we have two connectors linking elements in the base **B** with elements in the newly set up hypothetical space **H**. For example, in case *(1)*, element **b** in the base, associated with properties attributed to Gorbachev, including the name Gorbachev, is mapped via the identity connector **I** onto **b'** in space **H**. Element **a** in the base, corresponding to Henry, is also mapped onto **b'** in **H**, but via another connector **D**. Under the intended interpretation, connector **D** transfers some dispositional properties of **a** onto **b'** (e.g. Henry's aggressivity, or his political views), while identity **I** transfers (by default) the remaining properties of **b** onto **b'**, in particular relevant properties such as Gorbachev's 'situation': his role as head of state in the USSR, present political circumstances, etc.

Example *(2)* is formally similar to *(1)*, with *Bush* (element **a**) substituted for *Henry*. But the transfer function, call it connector **S**, works differently: this time the connexion between **a** and **b'** carries 'situational' properties from **a** to **b'**, while **b'** inherits its other properties, including dispositional, from **b** via connector **I**. Figuratively speaking, Gorbachev under the intended interpretation of *(2)* retains his character, along with other personal characteristics, but finds himself in the situation 'previously' occupied by Bush.

(3) is yet another case of the same formal construction, with a third type of connector, presumably identity, linking **a** (Henry) to **b'**, the counterpart, via **I**, of **b** (Lucy) in the base space **B**.

Diagrammatically, *(1)*, *(2)*, *(3)* give rise to construction *(4)*:

(4)
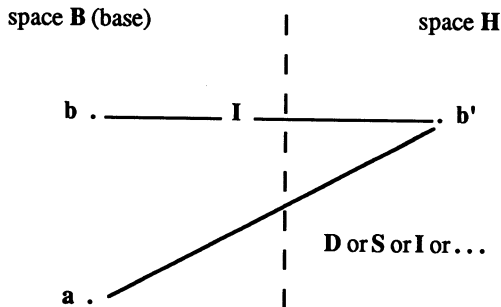
Typically, the construction of counterfactual spaces is used to produce 'real world' inferences. This will happen for examples like the ones above if further pragmatic principles or inferences apply. For example in the construction associated with *(1)*, the difference between **B** and **H** is that **b'**, but not its counterpart **b**, 'has' the property 'bomb Lithuania', and furthermore that **b'** (as opposed to **b**) has some relevant dispositions of Henry. For this to yield some 'real' properties of Henry, it must be further inferred that differences of disposition are behind differences in behavior (a plausible default assumption based on simplified socio-psychological models), and moreover that actions like 'bomb Lithuania' are symptomatic of particular dispositions (e.g. aggressive).

Notice that in the overt language expressions corresponding to such constructions, the copula **be** serves to signal the connector (as it does very generally for all connectors - cf. Fauconnier (1985), Chap.4). There is however no indication of what specific connector will be set up. Nor is there any indication of what pragmatic principles will apply.

2. Constraint satisfaction and space-building

Schematically, here are some of the operations that take place as meaning construction evolves:
- a central process of cognitive construction sets up domains, connects them in various ways, specifies relations satisfied within domains, links them to structured background knowledge (e.g. frames, schemata,..);
- this process is constrained at each stage of the construction by the previous configurational state, by pragmatic factors, and, last but not least, by the language expression that comes into the discourse at that point;
- further pragmatic principles apply in the exploitation of the constructed configuration, as exemplified above for example *(1)*.

In this section, we take a closer look at the grammatical constraint satisfaction part of the process. Consider once more examples like *(1), (2), (3)*. They have the general grammatical form *(5)*:

(5)   *IF*   NP$_1$ *be* NP$_2$,   Pro$_1$ Predicate

When a sentence having this form comes into the discourse, it will place constraints on the construction process by virtue of its grammatical structure and lexical content.

In order to outline the constraint process, I will use some notions from the mental space framework:[4]

- **space-builders**: a space-builder is a grammatical expression that either opens a new space or shifts focus to an existing space. Space-builders take on a variety of grammatical forms, such as prepositional phrases, adverbials, subject-verb complexes, conjunctions+clause. E.g. *in 1929, in that story, actually, in reality, in Susan's opinion, Susan believes___, Max hopes___, If it rains____, ...*
- **names** and **descriptions** (grammatically, noun phrases): names (*Max, Napoleon, NABISCO,...*), and descriptions (*the mailman, a vicious snake, some boys who were tired,...*) either set up new elements or point to existing elements in the discourse construction. They also associate such elements with properties (e.g. 'having the name Napoleon', 'being a boy', 'being tired'...).
- **presuppositional constructions**: some grammatical constructions[5] signal that an assignment of relations within a space is introduced in the presuppositional mode; this mode allows the relations to be transferred into neighboring spaces for the counterparts of the relevant elements.[6]
- **trans-spatial operators**: the copula (*be* in English), and other 'copulative' verbs, such as *become, remain*, may stand for connectors between spaces. (The general function of *be* is to stand for domain mappings; connection between spaces is a special case of this general function). Consider a grammatical structure of the form *NP$_1$ be NP$_2$*, where *NP$_1$* and *NP$_2$* are noun phrases, and identify elements a$_1$ and a$_2$ respectively, such that a$_1$ is in space X and a$_2$ is in space Y. Suppose F is the only connector linking spaces X and Y. Then the language expression *NP$_1$ be NP$_2$* will stipulate that a$_2$ in Y is the counterpart of a$_1$ in X via connector F, i.e.
$$a_2 = F(a_1)$$

- **identification of elements**:
   A crucial property of language and cognitive constructions is the following:

> **Identification:**
>
> If two elements **a** and **b** are linked by a connector **F** (b = F(a)), then element **b** can be identified by pointing to its counterpart **a**.

Linguistically, 'pointing to a' means to give a name or description of **a**. When this indirect identification procedure is used, we say that the element named or described, **a**, is the **trigger**, and that the element identified, **b**, is the **target**.[7]

- **identification path:** the above identification procedure can apply more than once. So if $b = F(a)$, and $c = G(b)$, and $d = H(c)$, element d at one extremity of the chain can be identified by pointing to element a at the other end of the chain. We say that there is an **identification path** from a to d.

- **role** and **value:** inside a space, there can also be connectors linking elements. One such connector is the role/value link V, such that if ' $a = V(r)$ ', element r is said to be a **role** for element a, and element a is said to be a value of that role. Typically (but not invariably), a role is pointed to by a description (e.g. *'the queen'*). This is a way of identifying the role (as in *'The queen selects the prime minister'*); but pointing to the role is also a way of identifying the corresponding value, by virtue of identification (identification path from r to a, via connector V).

Returning now to the general problem of constraint satisfaction, consider the grammatical form *(5)* once again. When a sentence of this form enters the discourse, constraints are placed on the ongoing construction. Such constraints include the following:

- conjunction *IF*: specifies that a hypothetical space H is open relative to the space B currently in focus; the relation specified by 'Predicate' must hold in that space;

- noun phrase NP_1: identifies some element (call this unknown 'x') in space B directly, or in space H indirectly by pointing to 'v' in B ;

- noun phrase NP_2: points to some element in B ('y') and identifies some element in H ('z');

- copula *be*: signals a connector 'F' that links the two elements identified by NP_1 and NP_2:

$$z = F(x)$$

(Notice that if x and z are in different spaces, F must then be a connector that operates across spaces; and if x and z are in the same space (necessarily H in this case), F will be the kind of connector that operates within spaces, e.g. role/value or metonymy.);

- pronoun Pro_1 must identify an element 'w' in space H that will satisfy the relation expressed by 'Predicate'. And since NP_1 is the grammatical antecedent of Pro_1, there must be an identification path linking 'v' or 'x' to 'w'.

What needs to be emphasized is that there is in principle an infinite number of subconfigurations compatible with such a set of constraints. This is because 1) the possible connectors do not constitute a finite set; 2) the identification paths from 'v' to 'x', from 'x' to 'w', or from 'y' to 'z' can in principle be as long as we like; 3) the possible prior stages of a construction, which serve as preconditions for a given step, are not finite in number either.

To see how all this works in practice, take the following instantiation of structure *(5)*, pointed out by Ch. Fillmore:

(6) If I were your father, I would spank you.

Fillmore noticed two very different interpretations for *(6)*: in one case (the 'lenient father' understanding), the speaker is suggesting that the father would be well advised to spank more; in the other (the 'brutal father' version), the speaker is pointing out to the child what the father is prone to do in certain circumstances (and the speaker may thereby be emphasizing his/her own leniency, or the child's luck, or the child's bad behavior,

or...). And there are in fact many more ways to read *(6)*. Here is how some of them come about.

First, *(6)* instantiates structure *(5)* in the following way:

NP_1 = *I*

Pro_1 = *I*

NP_2 = *your father*

Predicate = *spank you*

(I leave out the NP's corresponding to *you*, and the tense and mood markings)

Assume that space **H** is set up (by *if*), that *I* points to element **a** in the base space **B**, and that *your father* points to role **r** in space **B**.

In order to satisfy the constraints imposed by sentence *(6)* on the discourse construction, we must find the 'unknown' elements, 'x', 'y', 'z', 'w', and the unknown pragmatic connector 'F', specified in the constraint satisfaction schema for *(5)* outlined above.

Make the following simplifying assumptions: there was no prior discourse construction before the appearance of *(6)*; the only trans-spatial connectors to be considered are the transfer functions for counterfactuals studied in sec.1, **I**, **D**, and **S**; the only space-internal connector to be considered is the role/value link **V**.

Finally, for ease of presentation, let's give names to the participants, Mary to the speaker, Tommy to the child, Jack to the father.

Given all this, here are some of the ways to satisfy the grammatical constraints on the construction:

I) <u>dispositional interpretations</u>:

- role **r** has value **b** (Jack); **b** has counterpart **b'** in **H** (via the identity connector **I**); **a** (Mary) has counterpart **a'** in **H** (also via **I**);

- NP_1 (*I*) identifies **a** in space **B**;

- NP_2 (*your father*) triggers the identification path **r** ---V---->**b** ----I----> **b'**

- the connector 'F' signaled by the copula *be* is the dispositional transfer function **D**; it follows that **a** in the base is connected to **b'** in the counterfactual space by **D**:

$$b' = D(a)$$

- Pro_1 (*I*) triggers the identification path starting with **a**:

$$a \text{ ------- } D \text{ ------> } b'$$

It follows that **b'** is associated with the property indicated by 'Predicate'.

This construction can yield the 'lenient father' reading. Informally, **b'** in the counterfactual space inherits the speaker's relevant dispositions (via **D**), and the father's other characteristics (via **I**). I return below to the question of how this construction can be further elaborated in context to provide a full-blown interpretation.

II) situational interpretations:

Let the construction be exactly the same as in case I) except for the choice of connector 'F'. Instead of the dispositional transfer function, take 'F' to be the situational function S (cf. sec.1) that transfers situations across spaces. This time, the local situation L involving Tommy is, so to speak, transferred from Mary to Jack: **b'**, with b's properties inherited through **I**, is in situation L, and furthermore, for the same reason as in I), is associated with the property indicated by 'Predicate' (spanking Tommy). This makes the inference available (although not obligatory) that Jack in situation L behaves in the way indicated by 'Predicate', which is the source of the 'brutal father' understanding.

Diagrammatically, I) and II) correspond to *(7)*:

(7)
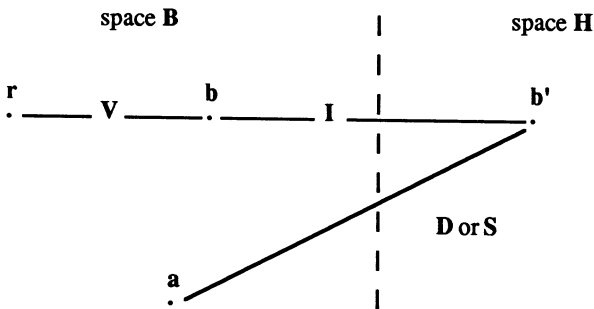
III) role interpretations:

Another way to satisfy the general constraints imposed by form *(5)* in the case of *(6)* is to have NP$_1$ trigger the identification path:[8]

$$a \text{------} I \text{------>} a'$$

and to have NP$_2$ trigger the path:[9]

$$r \text{------} I \text{------->} r'$$

With this choice of paths, the identification targets of NP$_1$ and NP$_2$ are both in **H**. The connector 'F' that must link them (constraint imposed by *be*) can be the role/value connector **V**, in which case:

$$a' = V(r')$$

With respect to the general constraint schema, what we have done here amounts to choosing **a** for 'v', **a'** for 'x', **r** for 'y', **r'** for 'z', and **V**$^{-1}$ for 'F'.

Informally, the effect of this construction is to shift the role 'Tommy's father' held by Jack to the speaker (I = Mary). This yields a class of understandings sharply different from I) and II): this time Jack does not come into the picture at all. In saying *(6)*, the speaker indicates that if she 'stood in the father relation' with respect to Tommy, then she would spank him. This does not even entail that Tommy actually has a father.

A diagrammatic representation of the construction for III) is *(8)*:

(8)
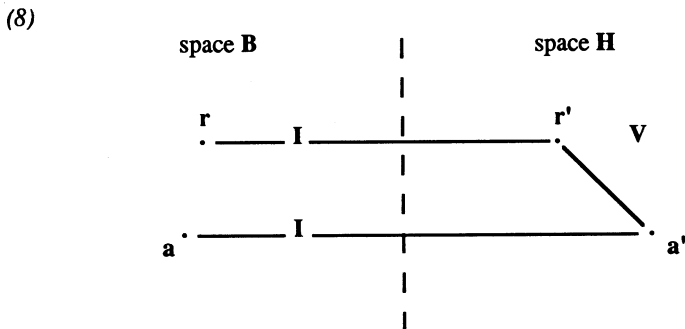
Constructions I), II), III) show how the same general grammatical constraints can be satisfied in superficially different ways. Other possibilities would emerge if we enriched our set of connectors or added more spaces to the configuration. For example, changing *spank* to *help* in *(6)*:

(6') If I were your father, I would help you.

suppose the context for *(6')* is the making of a movie and the casting has not been decided yet. The speaker is Kirk Douglas and the addressee Jane Fonda. The drama connector **T** (for 'theatre') becomes available, along with a space **M** for the movie relative to hypothetical space **H**. A host of new understandings arise by picking identification paths that use connector **T**, and target space **M**. Informal glosses of such understandings would be:

- Kirk tells Jane that if they were father and daughter in the movie, then in the movie, the father would help the daughter;

- Kirk tells Jane that if he were to play the part of Henry Fonda (Jane's real life father), then in the movie he would help whatever character she would be playing;

- Kirk tells Jane that if he were to play the part of Henry Fonda, then in the movie, the character Henry Fonda would help the character Jane Fonda (his daughter), who might of course be played by some actress other than Jane herself;

- Kirk tells Jane that if they were to be father and daughter in the movie, then he would help her off the set (in 'real life');

- Henry and Jane are both acting in the movie. Kirk tells Jane that if he (Kirk) were playing the part that Henry is actually playing, then the corresponding character in the movie would help the character played by Jane;

- Kirk tells Jane that if he were to play the part of Henry Fonda (Jane's real-life father), then he would help her in 'real life'.

- Kirk tells Jane that if he were her real father, then the character he plays in the movie would help the character she plays.

- Kirk tells Jane what he thinks the character in the movie ('her father') should do to the character she plays. (a dispositional construction vis-a-vis the character).

- Kirk tells Jane what the movie character would do in Kirk's real life situation. (a situational construction vis-a-vis the movie character).

The reader may justifiably find it tedious to go through these paraphrases of possible interpretations. The excuse for pointing them out is the following important and surprising theoretical point: the availability of such interpretations is <u>automatic</u>. It is a straightforward consequence of the **constraint satisfaction** conditions on meaning constructions: adding an extra connector and an extra space makes new identification paths available, and therefore increases the number of ways in which a construction can satisfy the constraints imposed by the linguistic expression.

To illustrate: the first interpretation in the above list corresponds to an identification path going from k ('Kirk') in B to k' in H (via connector I) and then to k'' in M (via connector T), and an identification path going from j ('Jane') in B to j' in H via I and then to j'' in M via T. The role r'' (father of j'') in M is connected to k'' via V. The third interpretation has a path going from r (father of j) to h ('Henry'), then to h' in H via I and to h'' in M also via I; another path goes from k to k' and then from k' to h'' via connector T; a third path goes from j to j' and then via I to j''.

To repeat a point often made in this regard: the possible configurations correspond to truth-conditionally distinct and clear-cut interpretations. In the particular context envisioned, sentence *(6')* is indeed at least twelve times ambiguous (not just vague, or incomplete). But it has only one structure, of the form *(5)*, yielding only one set of constraints on the configurations. The degree of ambiguity depends on the available spaces and connectors when the constraints apply.

3. <u>Pragmatic elaboration</u>

Although the construction process takes into account many factors of discourse and context that would traditionally be called pragmatic, the configurations obtained do not in themselves provide full interpretations. Full interpetations are obtained by further fleshing out of a properly pragmatic sort. Take for example the 'dispositional' construction corresponding to structure *(5)*, and applied to example *(7)*:

(7) If I were Nancy, I would give up.

This construction has dispositions of the speaker transferred (by **D**) to a counterpart of 'Nancy' in the counterfactual space. How can this yield real world inferences? The answer relies on simplified pragmatic default models that look informally something like this:

[Default model 1] Egocentric 'speaker does best' principle: the speaker's dispositions are the best and therefore lead to the most desirable actions or states of affairs.[10]

In *(7)*, we first have the inference in **H** (from the space construction) that 'someone' in Nancy's situation and with the speaker's dispositions would act in a certain way ('give up'). The default model provides the further inference that this course of action is recommended.

[Default model 2] Better than speaker is great: if speaker dispositions d_S lead to course of action a_S, and some other course of action a_n is better than a_S, then the dispositions leading to a_n must be 'better' than d_S, hence truly admirable.

In *(7)*, the space construction provides the same inference as before. But Default model [2] yields a different complete interpretation: since Nancy does better than 'give up', it is inferred that she has enviable dispositions leading to her course of action.

Notice then, that although [1] and [2] are both egocentric, [1] leads to an interpretation critical of Nancy, while [2] leads to an interpretation of praise. Other pragmatic interpretation strategies for *(7)* are available; for example, in a non-egocentric interpretation, the speaker could be taking Nancy's dispositions as the norm and deploring his or her (the speaker's) inadequacies. The choices among default models and interpretation strategies is a fascinating and hard question that will not be pursued here.[11]

4. Invisible meaning

We started out by observing the importance of Grice's work for highlighting the extent to which language underspecifies interpretation and understanding. The Gricean approach focuses on ways in which communication principles provide the means for 'filling in' what language leaves unspecified. But recent research taking into account the intermediate level of cognitive construction suggests that communication is not enough.

The position outlined above is more radical. Language expressions have no 'literal meaning' in the classical sense. What they do is impose constraints on space-building, leading to constructions which depend on context, but which also need further pragmatic elaboration to be complete.

The 'semantics' of a language expression is the set of constraints it imposes on cognitive constructions; this is a structural property, which is independent of context. The 'meaning' of an expression is something quite different. It arises, or rather some meaning arises, when a particular construction is performed and pragmatically elaborated.

It follows that 'implicature' cannot be viewed as merely a way to patch up basic meanings under pressure from communication principles, because there are no basic meanings for such a scheme to operate on. Instead, cognitive constructions require frames

and connectors which are not inherently linguistic, although language may code or highlight some of their characteristics. The frames and connectors bring with them rich (often non-monotonic) inference systems, and implicatures are typically part of such systems.[12]

It remains a formidable challenge, however, to find out how frames and connectors are chosen or retrieved in particular situations. The issue of relevance which Grice brought forcefully to the attention of linguists remains central in this regard.

---

[1] The linguistic and cognitive importance of the notion of role and/or frame is discussed in Fillmore (1982), Hofstadter, Clossman and Meredith (1982), Sweetser (1989), Sakahara (to appear a), Langacker (to appear), Fauconnier (1985, 1986). Goffman (1974) gives us a broad and insightful study of the sociological and contextual aspects of framing.

[2] This point has been stressed repeatedly in mental space research, and also in recent work on relevance, e.g. Sperber and Wilson (1986), and pragmatics (Recanati ms.).

[3] Connectors are studied in Maida (1984), Dinsmore (1989), Fauconnier (1985, 1990).

[4] Sakahara (to appear b) contains some recent research in this framework. Formalizations are proposed in Den (to appear), Dinsmore (1989).

[5] cf. Keenan (1971), McCawley (1981), Morgan (1973).

[6] cf. Fauconnier (1985), chap.3.

[7] Jackendoff (1975, 1983) shows how the equivalent of this principle applies analogously in the case of talk about pictures and about beliefs and how this explains the standard referential opacity puzzles. Nunberg (1978) applies the principle to metonymy and to pragmatic functions in general. The wide range of application of the Identification Principle to different kinds of domains and different kinds of connectors is studied in Fauconnier (1985, 1990).

[8] This identifies Mary's counterpart in space **H**.

[9] This identifies the role 'Tommy's father' in space **H**.

[10] Forman (1974) explains indirect speech acts in terms of a 'Speaker Knows Best' principle. Langacker (to appear) discusses the importance of subjectivity in accounting for grammatical phenomena.

[11] Recent research that studies in some detail the pragmatic aspects of space construction includes Rubba (1988), Lansing (1988), Kinsui and Takubo (to appear), Encrevé and de Fornel (1983), Encrevé (1988). Cicourel (1988) stresses the role of cultural and social factors commonly ignored in work on meaning.

[12] Especially revealing in this regard is Lakoff's (1987) study of idealized cognitive models (ICM's) and metonymic models.

Bibliography

Anscombre, J.-Cl. and O. Ducrot. 1983. *L'argumentation dans la langue*. Liège: Pierre Mardaga.

Cicourel, A. 1988. Formal Semantics, Pragmatics and Situated Meaning. *Center for Research on Language Newsletter* vol.2, no.3. UCSD.

Cole, P. and J. Morgan, eds. 1975. *Syntax and Semantics: Speech Acts*, vol. 3. New York: Academic Press.

Cornulier, B. de. 1983. *If* and the Presumption of Exhaustivity. *Journal of Pragmatics* 7.

Cornulier, B. de. 1984. Pour l'analyse minimaliste de certaines expressions de quantité. *Journal of Pragmatics* 8:5-6.

Cornulier, B. de. 1985. *Effets de sens*. Paris: Editions de Minuit.

Den, Y. to appear. Towards a Formalized Theory of Mental Spaces. In Sakahara ed. (to appear b).

Dinsmore, J. 1989. *Partitioned Representations*. Department of Computer Science, Southern Illinois University at Carbondale.

Dinsmore, J. to appear. Mental Spaces as a Theory of Representation and Language Understanding. In Sakahara ed. (to appear b).

Encrevé, P. 1988. C'est Reagan qui a coulé le billet vert. La dérivation généralisée. *Actes de la Recherche en Sciences Sociales* 71-72.

Encrevé, P. and M. de Fornel. 1983. Le sens en pratique. *Actes de la Recherche en Sciences Sociales* 46.

Fauconnier, G. 1975. Pragmatic Scales and Logical Structure. *Linguistic Inquiry* VI.3.

Fauconnier, G. 1985. *Mental Spaces*. Cambridge, Mass.: MIT Press.

Fauconnier, G. 1986. Roles and Connecting Paths. In Ch. Travis, ed. *Meaning and Interpretation*. Oxford: Blackwell.

Fauconnier, G. 1988. Quantifiers, Roles and Domains. In U. Eco, et al., eds., *Meaning and Mental Representation*. Bloomington, Ind.: Indiana University Press.

Fauconnier, G. to appear 1990. Domains and Connections. *Cognitive Linguistics* 1.1.

Fillmore, C. 1982. Frame Semantics. In Linguistic Society of Korea, ed., *Linguistics in the Morning Calm*. Seoul: Hanshin.

Forman, D. 1974. The Speaker Knows Best Principle. *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*.

403

Goffman, E. 1974. *Frame Analysis.* New York: Harper and Row.

Goodman, N. 1947. The Problem of Counterfactual Conditionals. *Journal of Philosophy* 44.

Green, G. 1975. How to Get People to Do Things with Words: The Whimperative Question. In Cole and Morgan (1975).

Grice, H.P. 1975. Logic and Conversation. In Cole and Morgan (1975). 45-58.

Hofstadter, D., G. Clossman and M. Meredith. 1982. Shakespeare's Plays Weren't Written by Him but by Someone Else of the Same Name: An Essay on Intensionality and Frame-based Knowledge Representation Systems. Bloomington, Ind.: Indiana University Linguistics Club.

Horn, L. 1972. *On the Semantic Properties of Logical Operators in English.* Ph.D. Dissertation, UCLA.

Jackendoff, R. 1975. On Belief Contexts. *Linguistic Inquiry* 6.1.

Jackendoff, R. 1983. *Semantics and Cognition.* Cambridge, Mass.: MIT Press.

Kay, P. and K. Zimmer. 1976. On the Semantics of Compounds and Genitives in English. In R. Underhill, ed., *Sixth California Linguistics Association Conference Proceedings.* San Diego: Campanile. 29-35.

Keenan, E. 1971. Two Kinds of Presupposition in Natural Language. In C. Fillmore and T. Langendoen, eds., *Studies in Linguistics Semantics.* New York: Holt Rinehart and Winston.

Kempson, R. 1986. Ambiguity and the Semantics-Pragmatics Distinction. In Ch. Travis ed., *Meaning and Interpretation.* Oxford: Blackwell.

Kinsui, S. and Y. Takubo. to appear. A Discourse Management Analysis of the Japanese Demonstrative Expressions. In Sakahara ed. (to appear b).

Lakoff, G. 1987. *Women, Fire and Dangerous Things.* University of Chicago Press.

Langacker, R. to appear. *Foundations of Cognitive Grammar*, vol.II.

Lansing, J. 1988. Delegated Points of View. ms. UCSD.

Maida, A. 1984. Belief Spaces: Foundations of a Computational Theory of Belief. ms. Dept. of Computer Science, Pennsylvania State University.

McCawley, J. 1981. *Everything that Linguists Have Always Wanted to Know about Logic.* Chicago: University of Chicago Press.

Morgan, J. 1973. *Presupposition and the Representation of Meaning: Prolegomena.* Ph.D. Dissertation, University of Chicago.

Nunberg, G. 1978. The Pragmatics of Reference. Bloomington, Ind.: Indiana University Linguistics Club.

Omori, A. 1988. Exploration into Mental Spaces: Where Do Linguistic Expressions Come From? *Journal of the English Linguistic Society of Japan* 5:130-149.

Recanati, F. 1988. The Pragmatics of What Is Said. ms. Groupe de Recherches Cognitives.

Ross, J. 1975. Where to Do Things with Words. In Cole and Morgan (1975).

Rubba, J. 1988. Mental Spaces and Ethnic Neighborhoods. ms. UCSD.

Ryder, M.-E. 1989. *Ordered Chaos: a Cognitive Model for the Interpretation of English Noun-Noun Compounds*. Ph.D. Dissertation, UCSD.

Sakahara, S. to appear a. Roles, the Particles ga/wa, 'eel' Sentences.

Sakahara, S., ed. to appear b. *Advances in Japanese Cognitive Science*. Vol.3. Tokyo: Koudansha Scientific.

Schmerling, S. 1971. A Note on Negative Polarity. *Papers in Linguistics* 4(1).

Seuren, P. 1985. *Discourse Semantics*. Oxford: Blackwell.

Sperber, D. and D. Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sweetser, E. 1989. Role and Individual Interpretations of Change Predicates. ms. UC Berkeley.

Travis, C. 1981. *The True and the False: The Domain of the Pragmatic*. Amsterdam: John Benjamins.

On the Foundations of Conversational Implicature

Richard E. Grandy
Rice University

Grice's work on logic and conversation is a famous episode of underground publication. The lectures circulated only in unauthorized manuscript form for five years after they were originally delivered in 1967, and were only published in fragmentary form until 1989. In this paper I will attempt to briefly discuss some philosophical background questions, mostly concerning the interesting question posed by Green (1987) "Why are there implicatures?"

Although Grice (1989:39-40) offered a number of tests for conversational implicatures--cancellability, detachability etcetera--he regarded these as only partial tests: "Indeed I very much doubt whether the features mentioned can be made to provide any such knockdown test, though I am sure that at least some of them are useful as providing a more or less strong prima facie case in favor of the presence of a conversational implicature" (43).

It is important therefore the useful test should not be confused with his official definition (1989:30-1): "A man who, by (in, when) saying (or making as if to say) that p has implicated that q, may be said to have conversationally implicated that q, provided that (1) he is to be presumed to be observing the conversational maxims, or at least the Cooperative Principle; (2) the supposition that he is aware that, or thinks that, q is required in order to make his saying or making as if to say p (or doing so in those terms) consistent with this presumption; and (3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition mentioned in (2) is required."

Two points are of special importance here. First, note that Grice requires that the hearer be able to work out the steps that lead to the conversational implicature. In the event that the requisite argument of the form described above cannot be provided "the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature" (Grice 1989:31). This means that an account of reasoning will eventually have to play a significant role in the analysis and assessment of conversational implicatures.

In evaluating Grice's claims about possible defenses of the classical logician's identification of the material conditional with *if-then* in English, the issue of distinguishing conversational from conventional implicatures is critical. Strawson (1986) has argued that a better account of *if-then* in English is obtained by seeing it as a 'ground consequent' relation between the antecedent and consequent.

The second point to emphasize is that the maxims or at least the Cooperative Principle must play a significant role in the argument. Some authors (Kasher 1976) have complained that very little conversation is cooperative. In fact, Grice was quite clear that the common ends that he envisioned might be quite vague or abstract and allowed many other motives to enter the conversational enterprise: "...we should recognize that within the dimension of voluntary exchanges (which are all that concern us) collaboration in achieving exchange of information or the institution of decisions may coexist with a high degree of diversity in the motivations underlying quite meager common objectives" (1989:369).

It may be useful to compare a stereotypical example of a situation which involves a high degree of cooperation, as when two persons wish to go to an airport to catch a plane with one driving and the other providing navigational directions, with situations that involve a degree of cooperation and a considerable degree of competition or disagreement. An intermediate case would be one in which both have a strong preference that they have dinner together at a mutually agreed upon restaurant, but have quite different preferences as to the kind of restaurant. Here there is a common goal but one that is not fully specific as to the way in which it is to be achieved and the main point of the conversation might be to specify that goal by providing relevant information or altering someone's perception of the alternatives or by changing their preferences.

In some cases, the cooperation with the interlocutor might be a superficial one imposed by an attempt at coordination with a third party. Professor Gumperz (this volume) presents two examples in his paper, one an instance of cross-examination in a courtroom, the other of a political debate. In both cases the protagonists have directly contrary interests and any gain by one is a loss by the other. Although in such cases the parties have directly opposed interests, they are constrained to carry on an interchange with a reasonable degree of cooperation and civility because each wishes to persuade the non-participating audience (jurors or listeners) of the correctness of their views and to do so requires at least some degree of cooperation in the conversation. Blatant failure to answer a question or striking the other participant with a chair is usually disastrous to achieving the persuasion of audience that is desired.

Some cases are more mixed. At present the major league baseball owners and players are engaged (through their intermediaries) in a conversation which is somewhat cooperative. The common goal is to reach a labor agreement which will be mutually agreeable and which will allow a season of games to be played this summer. If no games are played the owners make no profits and the players receive no salaries, so they are in agreement in wanting to avoid that mutually unpleasant outcome. However, within the range of alternative outcomes which involve some sharing of income, they are strictly antagonistic.

A second reason for the existence of implicatures has to do with features, some accidental and some not, of the lexicon. Some writers seem to suggest that being more informative requires a longer statement, and often it does. But there are also many cases of choice of lexical item which do not. The noun *cat* is both shorter and more informative than *mammal*. And thus the use of the latter very longer and less informative noun conversationally implicates that the speaker does not know more specifically which kind of mammal is in question.

This is related to the fundamental fact that language is general, i.e., nonspecific. When I ask for dessert, or an ice cream cone, the utterance has a semantic generality, but any ice cream that I buy or sell or consume must be totally specific as to flavor amount, taste, age and so on. Since almost all linguistic items are non-specific to some degree, the choice of one level of non-specificity rather than another always has potential significance.

The questions of with whom one is cooperating and for what goal is much more complex when conversations involving more than two persons are considered. In Rundquist (this volume) dinner conversations between families consisting of two parents and two children were analyzed for implicatures. In such situations there is an ongoing goal of the family unit to function as an appropriate unit, but the interpretations that the various individuals give to that phrase will differ considerably, just as the phrase "mutually agreeable labor contract" has rather different meanings for players and owners. Moreover the various subsets of the family, the two parents, the two children, the same/different sex parent and child dyads each have their own goals and these may conflict with those of other subunits. The entire topic of group and shared goals has received virtually no discussion, but appears to be a very significant topic for understanding the more complex realities of implicatures (Tuomela 1990).

Larry Horn (1984) has suggested a taxonomy of conversational principles into two categories following general principles of Zipf's. He uses labels Q and R to correspond to "...evoke Quantity ... and Relation while leaving open the extent to which my principles map onto these two maxims" (1984:13). His principles are:

(1a) The Q principle (Hearer-based):
      MAKE YOUR CONTRIBUTION SUFFICIENT...SAY AS MUCH AS
      YOU CAN
      (Given R)

(1b) The R Principle (Speaker based):
MAKE YOUR CONTRIBUTION NECESSARY...SAY NO MORE THAN
YOU MUST
(Given Q)

The project of analyzing the conversational maxims into more general or fundamental principles is certainly an interesting one and Grice was probably too concerned with making the maxims isomorphic to the Kantian categories rather than with their nature and relations. He remarks himself (1989:371-2) that the maxims presented are neither so coordinate nor so independent as the form of presentation suggested.

Nonetheless, I fear that the taxonomy that Horn is proposing may be misleading (I do not suggest that he is misled, but readers may be) because I think that it obscures two essential points. The first is that in considering the speaker based principle he suggests, or perhaps accepts Zipf's suggestion, that speakers need only be concerned with speaking truly and thus. But Grice emphasizes that speakers are performing social actions and have more concrete and important goals than maintaining their reputations as truth speakers, even though the latter is certainly of importance. Speakers typically want to achieve some specific effect on the hearer(s) and if this intended effect is specific, then not any old truth, or any undifferentiated truth, is adequate. If I want you to tell me how to get to the airport then I had best frame a direct question involving the airport or at least make a statement about the airport, airports generally or perhaps planes to catch, if I am intending to acquire the required information. In particular circumstances, of course, an apparently unrelated statement such as "It's 4:30" or even "It's Saturday" may suffice if the hearer (and context) provide the necessary connection.

And on the other side, the hearer does not want reams of information if a page will suffice. Cognitive scientists now regard it as a platitude that we are information processors, but probably do not sufficiently emphasize the extent to which our processing must consist of filtering. In the case of perception our evolution given sensory systems are admirably fitted for the purpose of eliminating a very high percentage of the available information at any time and focussing on the small percentage of information which is (generally) of the greatest importance. The mechanisms by which, for example, we manage to hear our name mentioned in one of many simultaneous conversations to which we were not paying attention is a familiar but still poorly-understood illustration.
Still I suspect, though I am quite uncertain how one might prove, that we are much less efficient at sorting through large quantities of orally presented information to find a small quantity of particular interest than we are at processing visually presented information. This may be a function of the nature of the task. When reading written material we are rather better at the task, but of course in that situation we have the option of reconsidering something that had been previously scanned or of stopping at a particular point, whereas in conversation although we can direct our attention we cannot (usually) direct the flow of the other's utterances.
Note that Grice's definition picks out the conversational implicatures from among the whole class of implicatures of a given utterance. Thus it is not a direct objection to his definition that he has not characterized the latter class, and some objections to his definition as too inclusive may be met by arguing that the problematic statements are not implicatures at all. For example, almost every statement requires for its truth the existence of physical objects and the speaker can expect hearers to realize this, if they reflect upon it, but it is not evident that this is an implicature of such utterances.

But as part of the larger project of applying the Gricean formulations to utterance we will eventually need a characterization of implicatures generally. In a previous discussion of Grice's work I suggested that at least one restriction would be that matters which are common ground between speaker and hearer would be ruled out as trivial

(Grandy 1989). This would rule out both general information commonly known, as well as matters that are idiosyncratic from a general point of view but common knowledge to this particular pair.

However, Robert Fogelin, in discussion, has produced a counterexample to my suggestion as follows: *A* is icing a cake and makes a mess of it. *B*, who has been observing, remarks "You've done a fine job!" Since the statement is obviously false the remark must be taken as sarcastic with the implicature that A has made a mess of things. But that is obvious to both *A* and *B*, and it is obvious to both that it is obvious to both, etc. Yet the implicature stands.

I do think that in general common ground is not material for implicature (at least if it is in the foreground of the conversation, if some matters are overlooked they could be routinely conversationally implicated). Perhaps the right corrective in the face of the counterexample is to assert the principle with the disclaimer that in cases of sarcasm, or obnoxiousness generally, anything is fair game for implicature.

A second kind of case to be ruled out concerns mutual knowledge of general information. An example adapted from Grice is:

A: Has Smith been dating anyone?
B: He has been visiting New York every weekend.

The implicature, according to Grice, is that Smith has been dating someone in New York. The problem is that there are other things which the theory seems to imply that *B* implicates. For example, it seems to imply that *B* implicates that New York is a place where people live (McCafferty 1987).

This particular implicature could be ruled out by excluding anything that was already common knowledge between the participants. Unfortunately, that exclusion does not suffice to keep out all unwanted candidates for implicature, because new knowledge may arise from the combination of background knowledge and the intended implicature. Imagine that the previous conversation takes place in Jamaica where *A*, *B*, and Smith all live. Since *A* and *B* both know that the only way to go to New York on the weekend is via Air Jamaica, *B's* response is consistent with his obeying the maxims, only if he believes that Smith flies Air Jamaica every weekend, and he can expect *A* to come to that realization. Moreover, it was not common knowledge before *B's* remark that Smith flies Air Jamaica every weekend. Thus, that proposition meets the conditions for being a conversational implicature (McCafferty 1987).

One way of responding to the criticisms would be to strengthen the third condition for conversational implicature and require not merely that the speaker think that the hearer will work out the supposition in the second clause, but that the speaker want the hearer to work out the supposition. In our last case, *B* may think, even expect, that *A* will work through to the conclusion that Smith flies Air Jamaica, but may be quite indifferent as to whether *A* does so. But *B* presumably does want *A* to come to the conclusion that Smith is dating someone.

Thus far I have been discussing some of the grounds from which implicatures arise. Another important and little explored area is the way in which conversational implicatures may disappear by being absorbed into the conventional. This is particularly relevant to some of the issues about the relation between formal logical languages and natural languages that are the most famous application of Grice's work. It is difficult to assess exactly what Grice's own position was with regard to the material conditional and natural language conditions.

Grice himself (1989:58-85) begins his lecture "Indicative conditionals" discussing the relation between *If p then q*, and in most discussions of his work no distinction is made between various forms which conditionals can take in English. Yet in a passage in

that same lecture (p.63) he distinguishes between the logical import of *If p, q* and *If p then q:*

> In fact, there seem to me to be quite a number of different forms of statement each of which has a good right to the title of conditional, and a number of which are quite ordinary or humdrum, such as "*if p, q,*" "*if p then q,*" "*unless p, q,*" and "*supposing p, (then) q,*" together with an indefinite multitude of further forms. The two forms which the strong theorist most signally fails to distinguish are "*if p, q*" and "*if p then q*"; and the strong theorist, therefore, also fails to differentiate between two distinct philosophical theses: (1) that the sense of "*if p, q*" is given by the material conditional, and (2) that the sense of "*if p then q*" is given by the material conditional. Thesis (1) seems to have a good chance of being correct, whereas thesis (2) seems to be plainly incorrect, since the meaning of "*if p, then q*" is little different from that of "*if p, in that case q,*" a linguistic form which has a much closer connection with argument than would attach to the linguistic form in which the word "then" does not appear.

More importantly even on the question what his position ultimately is regarding the relation between *if p, q* and the material conditional, one must decide whether to accept his disclaimer at the beginning of Lecture II (p.16):

> ...I shall confine myself to the dispute in its relation to the alleged diversions. I have, moreover, no intention of entering the fray on behalf of either contestant.

or the bold assertion that begins Lecture III (now Chapter 4, p.43):

> I am considering myself to have established, or at least put up a good case for supposing, that if any divergence exists between "if" and "QQ," it must be a divergence in sense (meaning, conventional force). I now aim to show, using the same material, that no such divergence exists.

Complicating the story further is Grice's discussion of <u>generalized conversational implicatures</u>. If a particular conversational implicature develops from situations which frequently recur, perhaps even are the normal situation, then an implicature that originally attaches separately to each occurrence may become more generally known. Thus if *if-then* is typically used in settings where the conversational implication of a connection between antecedent and consequent is present, it may become a matter of general knowledge and thus general expectation that the connection is present. Grice maintains, of course, that the distinction between what is said and what is conversationally implicated should be sustained even in these cases where the conversational implicature does not depend on specific features of the conversational exchange.

I have attempted to survey some of the sources of conversational implicature. These include the facts that we are rational social creatures who live in a complex world and require cooperation for many of our objections, that we have complex hierarchically organized lexicons as tools toward this end, that our information processing capabilities are highly limited. I have also indicated areas where further research is required to illuminate the nature of conversational implicature in general and Grice's own position in particular. Among the problems are a more thorough and explicit understanding of the role of reasoning and the potential working out of conversational implicatures, a better characterization of the process by which implicatures become generalized and a clearer grasp of the relation between general implicatures and meaning.

REFERENCES

Grandy, Richard E. 1989. On Grice on language. J Phil 86.514-25.

Green, Georgia M. 1987. Some remarks on why there is implicature. Studies in the Linguistic Sciences 17.77-92.

Grice, H. P. 1989. Studies in the way of words. Cambridge, MA: Harvard University Press.

Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. Meaning, form and use in context: Linguistic applications, ed. by Deborah Schiffrin, 11-42. Georgetown University Press, Washington, D.C.

Kasher, Asa. 1976. Conversational maxims and rationality. Language in focus, ed. by Asa Kasher, 197-216. Dordrecht Reidel.

McCafferty, Andrew S. 1987. Reasoning about implicature: A plan based approach. Unpublished doctoral dissertation. University of Pittsburgh.

Strawson, Peter. 1986. 'If and 'QQ'. Philosophical grounds of rationality: Intentions, categories, ends, ed. by Richard E. Grandy and Richard Warner, 229-42. Oxford: Oxford University Press.

Tuomela, Raimo. 1990. What are goals and joint goals? Theory and Decision 5.28:1-21.

The Universality of Gricean Interpretation

Georgia M. Green
University of Illinois

1. Introduction

The notion of implicature is now about 23 years old–over 33 if you trace it back to its origins in Grice's paper "Meaning", which emphasizes the importance, even for so-called literal (or "non-natural") meaning, of the speaker's intention in having her communicative intentions recognized. This is a notion that has proven critical to progress observed not only in the field of semantics, which it was originally intended to contribute to, but also in syntax, which it freed from bonds to non-semantic (i.e., non-truth-conditional) meaning, and in the study of language behavior generally (e.g., studies of rhetoric and politeness).

Nonetheless, perceptions of the domain of the principles that Grice described in "Logic and Conversation" are widely varied and quite disparate. My purpose is to show that, first of all, Grice intended a broad rather than a narrow interpretation of *conversation* relative to possible kinds of language use, and that, strictly speaking, the basis of implicature is the more general Cooperative Principle (CP) rather than the more specific maxims, which are neither culturally prescribed standards, nor mere customs, but more like default instantiations of the CP. I will also argue that Grice saw the relevance of the CP to language use as just the linguistic reflex of its general relevance. He saw it as governing all rational behavior. If it does, we should certainly find it universally applicable with regard to language use, and I know of no genuine counterexamples. Finally, it should figure in the interpretation of language generally, not just clever talk, and indeed, should figure in the interpretation of behavior generally, whether communicative or not. I will suggest that a case can be made that it does just that.

2. Conversation

It seems logical to begin with the narrowest possible interpretation of the domain Grice might have intended for his analysis of implicature. A very large number of writers who claim familiarity with the proposals of "Logic and Conversation" appear to believe that they pertain exclusively to oral communication, and in particular, oral communication that involves two or more individuals exchanging information, that is, what we call casual conversation or small talk. They appear to believe that Grice's analysis is not intended to apply to non-interactive modes of language use, e.g., to the production and interpretation of narrative, expository, persuasive or didactic discourse (including labels and instructions).

Thus, Sperber and Wilson (1986:261), for example, assert that "Grice's theory is put forward as a theory of 'conversation'", going on to say that it has "been taken as a more general theory of verbal communication, and Grice has done nothing to correct that interpretation." Pratt (1981) and Schauber and Spolsky (1986:116)

assume that the CP and the maxims do not cover situations in which only one person is talking.

Other writers seem to think that the scope of the theory is limited because it does NOT include small talk. Lyons (1977:593) appears to assume that the maxims do not apply to "utterances whose function is something other than that of augmenting the addressee's store of propositional knowledge," and so criticizes them for not relating to the "semantic information contained in everyday language-utterances" which he considers "social and expressive, rather than descriptive." Along similar lines, Stubbs (1983:101) expects different rules for social discourse ("small talk", "just chatting", "phatic communication") than for what he terms transactional discourse.

Despite the presence of the word *conversation* in the title of Lecture II, it is evident from examining the text that Grice was not referring exclusively to social conversation or small talk. As far as I can tell, Grice uses both *conversation* (seven tokens, excluding the title), and *talk exchange* (ten tokens) to refer to the same thing: the purposeful use of natural language.[1] In fact, in one place (1975:49) he equates conversation and communication, and refers to "the goals that are central to conversation/communication." Perhaps more significantly, of the 15 examples of natural language implicature that Grice gives (see Table I), only two strike me as representing small talk particularly (examples a, c about jail, and about whether Smith has a girlfriend), though several (e.g., b, h, j, k, l, and o) are pretty "gossipy", and only the three (examples b, d and l) that involve getting petrol, divulging someone's friend's address in the south of France, and changing the subject when someone referred to as an "old bag" enters the room necessarily involve oral as opposed to written discourse, and interactive as opposed to monologic discourse.

a. He hasn't been to prison yet.
b. There is a garage round the corner.
c. He has been paying a lot of visits to New York.
d. Somewhere in the south of France.
e. Mr. X's command of English is excellent and his attendance at tutorials has been regular.
f. Women are women. War is war.
g. P. It is certain that P. The evidence for P is Q.
h. X is a fine friend.
i. You are the cream in my coffee.
j. He was a little intoxicated.
k. She is probably deceiving him this evening.
l. The weather has been quite delightful. ...
m. I sought to tell my love, love that never told can be.
n. Peccavi.
o. Miss X produced a series of sounds that corresponded closely with the score of 'Home sweet home'.

TABLE I: Examples of particularized conversational implicature from
"Logic and Conversation"

Finally, in two places, Grice takes the trouble to single out casual conversation as just a special case of talk exchange.

> [The purpose or direction of a talk exchange] ... may be so indefinite as to leave very considerable latitude to the participants (as in a casual conversation). (1975:45)

> *In characteristic talk exchanges*, there is a common aim even if, *as in an over-the-wall chat*, it is a second-order one, namely, that each party should, for the time being, identify himself with the transitory conversational interests of the other. (Grice 1975:48; italics mine)

Likewise, despite the ubiquitous references to "talk exchanges", there is no reason to assume that Grice meant for his analysis to be an account of oral discourse to the exclusion of written communication as some writers (e.g. Pratt 1981) assume. Indeed, three or four of Grice's own examples (examples e, m, n, o in Table I), the letter of recommendation, Blake's line, the telegram from the British General, and the singing critique involve written language, and are described without any comment on this fact.

Grice's use of written language examples also indicates that it is a mistake to assume that he did not intend his analysis to apply to non-interactive discourse (such as narrative, exposition, and persuasive speech) on the grounds that they are not literally exchanges of talk. Written language and monologues would be artifacts of irrational behavior in the extreme if they did not presuppose the existence of some intended addressee or class of addressees, or at least that there would be some such.

3. The nature of the maxims

The central mechanisms which drive the Gricean inference engine are the notion, from the 1957 paper, that recognition of an agent's intention is part and parcel of identifying what act the agent performs, and the CP, the idea that rational language behavior is "such as is required, at the stage at which it occurs, by the accepted purpose or direction" of the enterprise in which the agent is engaged. Together these notions imply that speaker and hearer are constantly involved in the interpretation (usually not conscious) of what each other's goals must be in saying what they say. Grice described four categories of special cases of this principle (i.e., applications of it to particular kinds of requirements), and gave examples of their application in both linguistic and non-linguistic domains.

I speak quite deliberately in describing the "maxims" as just special cases of the CP. They are not (contra Lycan 1984:75) logical consequences (corollaries), because they don't follow as necessary consequences in all possible worlds. They are not additional stipulations. They are not necessarily an exhaustive list of special cases. Rather, they are just particular instantiations of ways of being cooperative; ALL OTHER THINGS BEING EQUAL, compliance with the CP involves conforming to all of them. When you can't conform to all of them, as Grice discusses, you do the best you can.

It is important to understand that the maxims do not *constitute* the CP either, as some writers seem to think. On the contrary, the CP is a very general principle which defines, depending on the values shared by participants, any number of maxims instantiating ways of conforming to it.[2] Thus, Sperber and Wilson's complaint [that Grice might need more than nine maxims] is beside the point:

> Are there just the nine maxims Grice mentioned, or might others be needed, as he suggested himself? It might be tempting to add a maxim every time a regularity has to be accounted for. However this would be entirely *ad hoc*. (Sperber and Wilson 1986:36)

Equally important, the maxims are not rules or norms that are taught or learned, as some writers appear to believe. Thus, Pratt (1981:11) refers to "Grice's rules", and Brown and Yule (1983:32) describe the maxims as "norms speakers operate with" and Blum-Kulka and Olshtain refer to them as "norms of conversation" (1986:174) and "normative rules of speech behavior" (1986:175). Allwood et al. (1977:37) refer to Grice (1975) as providing an account of "communicative norms which aim at making the exchange of information in a speech situation as effective as possible." Similarly, Ruhl (1989:96) refers to the maxims as "normative rules of appropriateness, adequacy, and relevance." Sperber and Wilson (1986:162) characterize Grice's principle and maxims as "norms which communicators and audience must know in order to communicate adequately,"[3] in contrast to their own principle, which they describe as a generalization about communication.[4]

Now, regardless of whether *norms* is understood as referring to prescriptive rules (cf. Brown and Yule 1983:32) or to customs (cf. Brown and Yule 1983:83), norms are not what Grice was referring to. It is really unfortunate that he articulated the maxims as imperatives, because there is nothing in Grice (1975) to suggest that he thought of them as statistical principles that people *tend* to conform to, or as ideals that people aspire to conform to. Rather, they are described as governing the communicative acts of *all* (sane) communicators.

That the maxims are just obvious ways of BEING cooperative, and not rules we have to learn, suggests that they may only come to our attention when we encounter speech which is hard to reconcile with the assumption that they are being observed, and this seems to be the case. Sperber and Wilson's Principle of Relevance can be seen as not being effectively distinct from the claim, consistent with the CP, that the maxim of relevance is always assumed to be observed.

Perhaps I can illustrate the tenacity with which we assume that speakers observe the CP, and in particular the maxim of relevance, by examining what we might make of an apparently irrelevant speech act. Suppose a stranger were to approach me in a public place, and after introducing himself, ask the time. I would surely think this very strange, and wonder why he introduced himself.

1. Excuse me, I'm Sterling Ryznich. Can you tell me the time?

Why did he think I would care who he was if all he wanted was the time? It is only when we come across an utterance whose relevance is not easily inferred that we notice that we expect every utterance in a discourse to be relevant to some participant's goal which is at least mutually accessible, if not truly shared. If the hearer assumes that the speaker is abiding by the CP,[5] he must adopt a strategy of interpreting the speaker's behavior as conforming to the maxims, and considering what propositions must be assumed in order to make that behavior patently in conformity with the CP and the maxims.

Since the speaker (as a speaker-hearer of a natural language spoken in a society) expects the hearer to adopt this strategy for interpreting speech behavior, the speaker is free to exploit this fact, and speak in such a way that his behavior must be interpreted according to it. If the speaker's remark seems irrelevant, the hearer will seek to construct a sequence of inferences which make it relevant to some assumed goal.

In the case of (1), we have two choices: either assume that the introduction is truly irrelevant to any purpose of speaker or hearer, in which case we should consider the stranger truly deranged, or assume that there is some goal to which the stranger's identity is relevant. Maybe it is a goal of his: maybe he wants to use me as an alibi (I'd certainly remember bizarre behavior like uttering (1)). Maybe he is trying to terrorize me psychologically, and just wants me to worry about why he said what he said. On the other hand, maybe it is relevant to some goal he imputes to me. Maybe he (mistakenly) thinks I am someone who was supposed to meet him. Maybe he thinks he is such a celebrity that even though he never expects to see me again, I would want to know whom I had befriended. Maybe he just wants to know the time, but believes that the injunction against speaking to strangers will inhibit me from answering him unless he introduces himself, making him no longer a stranger. And so on. Only our imagination limits the goals, and beliefs about my goals, that we might attribute to the speaker. The point is that if we reject the assumption that the speaker is irrational, then we must at the very least assume that there is some goal to which his introduction is relevant in the context of asking the time, even if we can't imagine what it is.

4. Purposes and rationality

One further point seems to require discussion here. What Grice intended to refer to by *purpose* and *goal* in "Logic and Conversation" is apparently even less clear than the intended referent of *conversation* and *talk exchange*. Others' strategy has apparently been to take a fairly literal approach to both questions, investing Grice's numerous hedges and apologies (e.g., pp. 47, 49) with a great deal of significance, and arriving, logically, at very narrow interpretations. Thus, Sperber and Wilson (and Pratt and many others) seem to think that the purposes of participants in talk exchanges which Grice refers to must be mutual, cooperative purposes, and of a knowledge-oriented nature, like understanding the economy of Rumania, or figuring out how to put together a pump.

Grice assumes that communication must have 'a common purpose or set of purposes, or at least a mutually accepted direction' (Grice 1975:45) over and above the aim of achieving successful communication. We do not mean to deny that this is very often true, particularly in conversation. In a talk exchange, a seminar or a book, there may well be a mutually manifest purpose or direction. (Sperber and Wilson 1986:161-162)

Now, I have read "Logic and Conversation" a number of times over the last 23 years. Since I have been unable to arrive at a coherent, consistent literal interpretation, my strategy has been to assume that Grice's ability to say clearly what he meant just did not match his vision,[6] and therefore, to interpret problematic terms in such a way as to arrive at a consistent and explanatory theory. To me, this beats accepting either of the alternative conclusions: that Grice's goal (and thus the scope of his analysis) must be very narrow, or that he was short-sighted (to put it politely), and simply did not think through the consequences of the propositions constituting his analysis. I do not claim credit for the interpretation I support here, since I think it is at least implicit in what he wrote, and since I am sure I have been influenced in its favor by others,[7] but if it should turn out to be not what Grice meant at all, I should still prefer it.

Grice indicates at several points that he sees discourse as purposive behavior. That is, he presumes that participants have goals in participating (apparently since otherwise they wouldn't be participating). This is the gist of the first passage referring to purposes:

each participant recognizes in [talk exchanges], to some extent, a common purpose or set of purposes, or at least a mutually accepted direction. (Grice 1975:45)

Grice is very vague about these purposes: how many there are, how shared they have to be. With twenty years of hindsight, and close attention to this question, we can say that the purposes are first of all not unique. Conversants typically have hierarchically embedded goals. Taking the most transparent sort of case, Martha may say "X" to George, with the immediate goal that George understand that Martha has said "X", and meant some transparently derivable proposition p by it, so that George will believe that Martha believes that p is true, so that George, respecting Martha's opinion, will come to believe p himself. Presumably Martha has a reason for wanting George to believe p (another goal), and probably reasons (and goals) behind that. If we were to assume that only one of these was "Martha's goal", I don't know what grounds we could use to decide which one it was, that is, which was the unique correct level of analysis.

Second, goals are not so much shared or mutual, as they are mutually modelled (Cohen and Perrault 1979, Cohen and Levesque 1980, Green 1982, Appelt 1985, Cohen and Levesque 1990, Perrault 1990): for George to understand Martha's utterance of "X" to George, George must have beliefs about Martha which include Martha's purpose in uttering "X" to George, which in turn includes Martha's model of George, including George's model of Martha, etc.[8] Grice's comment (1975:48) that

"in characteristic talk-exchanges, there is a common aim even if...it is a second-order one, namely, that each party should, for the time being, identify himself with the transitory conversational interests of the other" can be interpreted along these lines. The idea that participants will at least temporarily identify with each other's interests, i.e., infer what each other is trying to do, is what allows quarrels, monologues, etc. (pace Pratt and Schauber and Spolsky[9]) to be included in the talk exchanges that the CP governs. The participants may have different values and agendas, but given Grice's (1957) characterization of conventional meaning, for any communication to occur, for each partici-pant to understand what the other meant, each must make assumptions about the other's goals, at least the low-level communicative goals. This is the sense in which participants recognize a "common goal". When the assumptions participants make about each other's goals are incorrect, and this affects non-trivial beliefs about each other, we say they are "talking at cross-purposes", precisely because of the mismatch between actual goals and attributed goals.

Finally, and I think Grice's recognition of this is clear from the continuation of the passage on p. 45, and from remarks three pages later, each individual in a talk exchange may be directing her behavior towards a multitude of goals (or purposes), and these may be independent of each other, or hierarchically linked.

> This purpose or direction may be fixed from the start (e.g., by an initial proposal of a question for discussion), or it may evolve during the exchange; it may be fairly definite, or it may be so indefinite as to leave very considerable latitude to the participants (as in a casual conversation). (1975:45)

> The participants have some common immediate aim, like getting a car mended; their ultimate aims may, of course, be independent and even in conflict– (1975:48)

Thus, S may have as a goal to get H not to vote for Sen. Foghorn, but she may also intend to maintain good social relations with H, and therefore intend to flatter H (or boost his positive face) in order to get H to believe that not voting for Foghorn was his own idea.

Grice's references to discourse as purposive behavior, are regularly linked with references to the rationality of human beings, and it is this fact, along with the analogues in three separate places on pp. 47 and 48 to non-linguistic behavior, that invites us to understand the CP as something that governs not just discourse, but rational intentional behavior generally.[10]

> Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes... at each stage, SOME possible conversational moves are excluded as conversationally unsuitable. (1975:45)

> one of my avowed aims is to see talking as a special case or variety of purposive, indeed rational, behavior (1975:47)

> one feels that the talker who is irrelevant or obscure has primarily let down not his audience but himself. (1975:48-49)

Insofar as the first and last passages are equally true of agents and interactants generally, not just talkers, they support the interpretation of the CP and the attendant maxims as descriptive of and relevant for interactive behavior generally, not just discourse.

The discussion of example (1) indicated that interpreting the communicative behavior of other human beings as intentional, and relevant, necessary, and sufficient for the achievement of a presumed goal seems to be unavoidable—maybe we are born to do it. This is equally true of behavior that isn't intended as communicative. For example, Morgan (1978) described a situation where someone is standing at a door, gripping a key that is inserted in a lock, and making partial rotary motions with his wrist and forearm. It is very difficult to describe this without prejudicing the interpretation, but that only strengthens my point: no matter how I describe it, if we imagine it, we impute some purpose to the person, and it involves the key and the lock—probably using the key to lock or unlock the door, but it could be something less usual, like an isometric exercise. My point is that we do not seem to entertain the idea that he might be doing it for no reason. That alternative, along with the possibility that he isn't even doing it, that it's just happening to him, is one that I think we are very reluctant to accept without any independent support, such as knowledge that people do that sort of thing as a nervous habit, like drumming their fingers on a hard surface, or playing with their hair.

The understanding of social meanings in casual conversation may work similarly. When Martha asks how George is, or states the obvious to make social conversation, it isn't so much the propositional content of what Martha says that informs George and contributes to a mutually accessible goal. Rather, it is the fact that Martha has gone to the trouble to ask the question, or evoke the happy or commiserable thought, that is interpreted as intentional behavior. Interpreting it as intentional, George unavoidably tries to infer why it was intended (e.g., to cause him to believe that Martha cares enough to go to the trouble to make small talk). As with other sorts of meaning, such inferences about intentions are intended to be made. Thus, the sharing of "social and expressive" information in small talk seems to be governed by the same general principles of interpretation that govern the interpretation of other intentional and communicative acts.

I hope I have made it clear that "cooperative" in the sense of the Cooperative Principle does not entail accepting your interlocutor's goals as your own and helping her to achieve them. Rather, it is most usefully understood as meaning no more—and no less—than 'willing to try to understand the interaction from the other participants' point of view', i.e., to try to understand what their goals and assumptions must be.

5. The geographical generality of Grice's principles

In talking about the observance of the CP as rational, Grice (1975: 47, 49) implies that he takes it and the maxims to represent values universally assumed in human society. Grice does not explicitly claim universality for the CP and the maxims, but if they are to explain discourse phenomena as a function of rationality, they have to be universal (barring societies of irrational beings), and hence potentially a consequence of some property of human nature or human society, and not just of familiar English-speaking cultures. Consequently, insofar as the maxims are only instantiations in a context of the CP, and not corollaries, discovering that one of the maxims was not universal would not invalidate claims that the Cooperative Principle was universal.

However, attempts to show that one or other of the maxims is not valid in some society usually end up showing something a little different, namely that conforming to the maxims is constrained by cultural values, such as deferring to superiors, being indirect or noncommittal (to protect one's "face" (or cover one's ass)), belonging to an information elite, protecting the "negative face" of others (e.g., by refraining from making personal references or otherwise imposing, and from assigning blame, etc.). Thus, being cooperative in making one's contribution "such as is required at the stage at which it occurs" to accomplish one's goal(s) may involve following other principles as well, such as:

A. Assigning responsibility for a state of affairs that could be construed as undesirable is counterproductive.
B. Making the addressee disinclined to cooperate (e.g., by making the addressee uncomfortable) is counterproductive.
C. Maintaining social advantage is useful.

What counts as an undesirable state of affairs, as making someone uncomfortable, and as constituting social advantage, are of course defined relative to a particular culture.

Thus, cultures can be expected to differ in the values they assign to various ways of being cooperative, and matters of avoiding offense and losing face can be expected to interact with these, since the behaviors that count as offensive and as contributing to status are well-known to be defined differently in different cultures. But it would astonish me to find a culture in which Grice's maxims were not routinely observed, and required for the interpretation of communicative intentions, and all other things being equal, routinely exploited to create implicature.

6. The behavioral generality of the principles

If the CP and the maxims are understood as principles of rational behavior–or as principles defining rational behavior, one would expect to find them involved in interactions of all sorts, not just "clever" uses of language, like the gossipy or catty uses Grice cites (e.g., examples a, d, e, l, n, o in Table I). To begin with, the case

has been made (by both Nunberg (1978) and Sperber and Wilson (1986)) that the literal interpretation of linguistic expressions in general involves the same mechanisms as conversational implicature, though Grice came close to recognizing this in his discussion of generalized conversational implicature as involved in examples like (2) (cf. Horn (1984)).

2a.   X is meeting a woman this evening.
2b.   X broke a finger yesterday.

Nunberg (1978) argued (cf. also discussion in Green (1989)) that the maxims and the CP are involved in a process paralleling conversational implicature in determining the reference in context of uses of ordinary nouns and verbs such as *newspaper* or *run*, and proper names, as in (3-5).

3.   Rupert bought another newspaper.
4a.   Ronnie is running.
4b.   The engine is running.
4c.   The program is running.
4d.   The electricity is running.
5a.   The population of Chicago is over 5 million.
5b.   Chicago beat Dallas 44-0.

Nunberg argues that such terms are not in general lexically ambiguous, but vague, and their interpretation in context is a matter of inferring first, what class of referents the speaker takes to be normal for that term in that speech situation (i.e., making assumptions about the hearer's beliefs about the world), and second, what derivative function relates the referent she intends to that class. Understanding the contextually indicated reference of metaphors and other extensions of "literal" meaning, for example, a *spark* of anger, to *sail* a report across a desk, follows as a particular subcase, under Nunberg's analysis (Nunberg 1988).

Whether or not Grice realized the variety of examples that could be explained by analyses along the lines he suggested for *and* and *or* is not really important, and whether one should characterize interpretation of non-obvious purposive behavior generally as implicature seems to me an arid question of nomenclature, not a question of analysis.

In any case, a further consequence of the CP, and of the maxim "Be relevant" in particular, that Grice never discussed, is that it provides the basis for a natural account of the problem of the coherence of texts, that is, of elucidating what it is that makes (6) a plausible discourse, in contrast to (7), which appears to be just a string of sentences related only by temporal or spatial order.

6. The following days were unlike any that had gone before. There wasn't a man on the ranch who didn't know of Saturday's race and the conditions under which it would be run. They gave any excuses to get near the black stallion's corral. [Walter Farley, *The Black Stallion*, p. 199. New York: Random House (1953)]

7. The sun climbed higher, and with its ascent the desert changed. There was nothing Lucy liked so much as the smell and feel of fur. One evening, after dark, she crept away and tried to open the first gate, but swing and tug as she might she could not budge the pin.[11]

If we assume that the underlying organizational principle for discourse is simply the assumption that the sequence of sentences at issue is produced by an individual in order to achieve some goal in accordance with the CP, then it follows, given no indication to the contrary, that each sentence is uttered in the execution of some plan to achieve that goal and conforms to the maxims that instantiate the CP, and is thus intended to say something necessary, true, and relevant to accomplishing that goal. Then, a coherent text is one where the interpreter can readily reconstruct the speaker's plan with reasonable certainty, by inferring relations among the sentences, and their individual relations to the various subgoals in the inferred plan for the enterprise or enterprises understood to be at hand.

Coherence, in this approach, depends not on properties of the text components themselves, either individually or in relation to each other, but rather, on how great an effort is required to construct a reasonable plan to attribute to the text-producer in producing the text. This in turn depends on how hard or easy it is to take each sentence as representing a true, necessary, and relevant contribution to that plan.

Some writers (cf. Lakoff 1973, Leech 1983, Brown and Levinson 1987) have interpreted the CP and the attendant maxims as constituting a theory of informative behavior, which they see as generally contrasting and/or conflicting with principles governing politeness, apparently assuming that the goals of being polite and being informative tend to conflict. Grice himself implies that "Be polite" ranks as a maxim of the CP along with Quantity, Relevance, etc.:

There are, of course, all sorts of other maxims (aesthetic, social, or moral in character), such as 'Be polite', that are also normally observed by participants in talk exchanges, and these may also generate nonconventional implicatures. (1975:47)

The CP would seem to entail politeness whenever the addressee's positive feelings toward the speaker are essential to the speaker's accomplishing her goals. Brown and Levinson (1987) recognize a number of ways in which what they call Grice's "conversational maxims" (Quantity, Quality, Relevance and Manner) are exploited in the service of politeness strategies.

I want to suggest that more generally, a large measure of what makes any linguistic act polite must be attributed to the assumption that participants are acting in accordance with the CP and the various maxims. I follow Brown and Levinson's practice in using *politeness* to refer to whatever means are employed to display con-

sideration for one's addressee's feelings (or face), regardless of the social distance between the speaker and addressee. Insofar as being polite, on this interpretation, involves a wide variety of strategies depending on estimates of the social distance between speaker and addressee, the relative social status of speaker and addressee, and the extent to which the act contemplated is considered to be an imposition in the culture which speaker and addressee are members of, when I refer to "a politeness maxim", I will have in mind something like maxim B above.

Politeness, according to Brown and Levinson (1978, 1987) is trade in a commodity they call face. FACE is defined as consisting of two components, first, the freedom to act unimpeded (negative face) and second, the satisfaction of having one's values approved of (positive face) (Brown and Levinson 1978:67). To engage in normal interaction is to risk losing face, since being asked to do something implies disapproval of your not having done it already, and puts pressure on you to act a certain way. Similarly, telling someone something implies that they did not know it already, and drawing attention to that ignorance threatens the face of the addressee, since it may amount to implicit criticism of their values and choices. Just talking to someone constitutes something of an imposition on their time.

Since face is defined in terms of approval by others, and absence of interference from others, maintaining face requires the cooperation of others, at least insofar as it requires that they show approval and refrain from acts that would constrain one's own activity. Since threatening the face of the addressee is likely to make him less willing to cooperate with the speaker in maintaining her face, and in any other goals he imputes to her, one would expect all enterprises which require cooperative interactions to involve the assumption that interactants are guided by the CP. Thus, interactants trade in face, paying face whenever in the course of accomplishing their goals, they must perform a face-threatening act such as conveying information that may be interpreted as implying a criticism, or making a request for goods or services, or even for the time that it takes to share information.

The supposed conflict between the "conversational" maxims and a politeness maxim is assumed to be obvious insofar as it is assumed to be common knowledge that it is impolite to be direct about certain things: making critical remarks, asking personal questions, bossing people about. But observed conflicts are between two kinds of goals: those that have to do directly with sharing information and inducing action, and those that have to do with preserving or enhancing the social status quo, by preserving or enhancing one's own "face" or that of one's addressee. Conflict between these goals is not a necessary consequence of anything; learning something from someone does not ENTAIL feeling threatened by that knowledge. Listening to someone is not a waste of time if you profit by it. Indeed, attaining goals that involve sharing information or inducing action regularly depends on success in keeping the addressee happy and cooperative by maintaining his face. Consequently, making one's contribution such as is required at the stage at which it occurs will involve whatever is necessary to secure the cooperation of the addressee (cf. Maxim B). Frequently this will involve satisfying face wants, and avoiding threats to the face of the addressee. Thus, Maxim B, and such submaxims as "Don't impose", "Offer options", and "Make the addressee feel good",[12] follow from the CP in the same

way that "Be brief" and "Be orderly" do. They are additional ways of making your contribution "such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice 1975:45). At the same time, politeness displays convey information about the speaker's beliefs and wants concerning the relationship between speaker and addressee, and consequently, implicature is a key component in a number of politeness strategies, not just the move–relatively drastic in Brown and Levinson's hierarchy (Brown and Levinson 1978:65)–of going off-record by implicating rather than speaking in a more direct fashion. The relation between being polite and being informative is one of symbiosis regulated by the CP, not one of antagonism between the CP and a politeness maxim on the order of "Encourage cooperativeness".

We can examine a couple of cases as illustration. A common positive politeness strategy (making the addressee feel good) involves the use of special deferential forms, such as formal second person pronouns in German and Spanish, and deferential affixes on nouns and verbs, as in Japanese and Korean. By convention, the use of deferential forms indicates the speaker's respect for the addressee, referent, or bystander; it is customary to use them when addressing a person to whom one is expected to show deference, for example, teachers, parents, grandparents. Consequently, they are typically employed 1) to mitigate face threat by feeding the addressee's ego, by explicitly acknowledging her superior status, or 2) to minimize face threat by implying that since the addressee is so superior in status and authority to the speaker, the speaker is not in a position to demand or force any action or reaction from the addressee. Here, it is particularly implicatures from the formal, possibly comparatively obscure manner of speech that create the intended positive politeness effect. As one might expect, speakers of Korean and Japanese further exploit this rich system in sarcasm and in switching levels of honorification to implicate that the relationship between the speaker and the addressee has changed, or that it should change to the relation implied by the honorific morphemes. Similarly, in English- speaking cultures, strangers who want to have the privileges of an intimate relationship (e.g. insurance salesmen and telephone solicitors) often use positive politeness tactics to imply that that relation already exists, asking personal questions about the addressee's family and state of mind (*Good evening. How are you tonight? And how is Mr. Green?*). Often telephone solicitors don't identify themselves at the beginning of the interaction, and address you by your first name, speaking as if you should know who they are and reciprocate with intimate, positive politeness behavior.

Avoiding imposing by employing hedges and euphemism and conventional indirect speech also amounts to performing a linguistic act in a particular way. Thus it also involves an implicature, according to the maxim of manner, from the fact that the speaker took the trouble to go through the motions of attempting to protect the addressee's face. Furthermore, as Lakoff (1973:305) pointed out, insofar as "Don't impose" entails "Don't waste time", being polite according to this strategy entails being considerate by conforming to all of the so-called informativeness maxims.

Finally, a third strategy consists of going off-record and implicating even more obliquely the opinions and wants that the speaker would like to convey without the

threat to the addressee's face that direct expression might imply. Going off-record involves just hinting at what you want or think, for example, saying *I wish I didn't have to walk home* instead of asking for a ride, or saying *That style is also available in a size 14* instead of saying *You'll split the seams of that size 12 if you try to put it on.* Both the politeness aspect of hints and the communicative aspect work the same way as such politeness formulae as hedges: the content intended to be communicated must be inferred via the assumption that what was said is relevant to salient concerns; its politeness value is a function of the recognition of the fact that direct expression was intentionally avoided.

My point has been that the basic principles of Gricean inference are universal, and far from being specific to particular kinds of language use, are not even specific to language, but rather are part of the human condition; people of all cultures cannot help interpreting each other's actions in terms of a background assumption that the actions are rational according to the CP and whatever "maxims" follow from it IN THAT CONTEXT. I have tried to show that this does not preclude cultural differences on a grand scale. I claim that such differences are not random or arbitrary, but follow in a principled way from cultural differences in values. Because of the nature of the basic principles, it follows that they develop (as opposed to being learned) in all individuals who understand the rest of their species to be goal-directed planning organisms that act autonomously. Values, of course, have to be learned. Sometimes they have to be taught. And there is apparently a lot of individual variation in how well they are learned. Individuals also differ in· their ability to correctly classify situations. And only a few may excel in developing a repertoire of techniques for balancing conflicts among goals and values–this is a fine art.

I would like to conclude by suggesting that one of the major values of implicature and exploitation of the CP in general[13] is that it allows us to mean what we say at a number of different levels, and that as speakers and hearers, it is a source of enjoyment to do so. The typical implicature (I exclude sarcasm as a very special case) involves a more or less literal meaning of what is said. When you say *It's cold in here* to get someone to do something about it, you do mean that it's cold where you are. You wouldn't say it if you thought it was false. The typical implicature also involves a chain of inferences leading to the intended understanding; in this case: 'I'd like you to do something to make it warmer'. And typically, the fact that there is a difference between what was said literally and what was intended to be conveyed is also significant, and intended to be understood: 'I have spoken in this oblique fashion out of consideration for your feelings'–an inference involving the manner maxim.

Verbal art is the most obvious example–puns and higher rhetorical figures such as metaphor. The sign that says "Centipede Crossing–100 feet" is funny/satisfying/irritating because it is literally true on two interpretations. It is deliberately ambiguous, and you're supposed to recognize that it is and that you are supposed to recognize **that**.

Political art, both in the sense of persuasive rhetoric and in the sense of diplomacy, including being diplomatic generally, involves saying things in such a way

that they will be interpreted in different–favorable–ways by different groups in the same audience (Myers 1984). Yassir Arafat's ambiguous denunciation of terrorism in 1988 is a good example of this, except that it didn't fool anybody.[14] Apparently, the *art* of diplomacy involves the art of minimizing the depth to which hearers will search for the intentions behind intentions.

My final example involves graphic art, in the sense of communicative effects of the way something is represented in visual rather than verbal terms. In one of the euphoric demonstrations following the opening of the borders between East and West Germany, the following sign was displayed.

Nothing I can do with my vocal apparatus or my typewriter can reproduce the essence of this sign. To understand its message, you have to notice the difference between the token of 'reunification' (*Wiedervereinigung*) that is in green and red roman capitals and the token in brown Fraktur. The sign-bearers are opposed to (*gegen*) old-style, Third Reichian, brownshirt reunification. They favor a modern-style reunification, involving the environmentalist Greens and the communists (the "Reds")–or on another reading, the Social Democrats, whose color is red. You have to recognize the different typestyles and colors, and realize that you are intended to recognize them, and attribute significance to them, inferring reasons why each is the way it is, and not some other way, and assume that you are intended to infer just those reasons, and intended to recognize that you are so intended, etc. Precisely the logic of "Logic and Conversation", and of Grice (1957) as well.

The spirit of Grice's theory is that behavior, communicative and otherwise, is goal-directed and is correctly interpreted only by acknowledging that it is intended to be so interpreted. The fact that Grice did not clearly articulate its breadth does not diminish its explanatory power.

FOOTNOTES

[1]Three other expressions, *natural language* and *natural speech* or *natural counterparts* occur in the introductory pages. They all seem to refer to language as a system, as opposed to utterances in a language, which is what I take *talk exchange* and *conversation* to refer to.

[2] Stubbs (1983:95) may recognize this in referring to the maxims as "maxims of co-operativeness", but he does not discuss how the individual maxims he invokes might be related to the more general principle.

[3]In general, it seems to me, Sperber and Wilson's criticisms (see especially p. 37) are based on the false assumptions that communication is always perfect and that its perfection is what must be explained. Although they reject the code model of communication embodied by the conduit metaphor exposed by Reddy (1979), they do not consider his toolmakers' metaphor, which treats fully successful communication as comparatively rare, and generally the result of hard work by both speaker and addressee. They retain the conduit metaphor premise that if communication is unsuccessful, blame lies squarely at the feet (mouth?) of the speaker, in saying (1986:251) that "the first interpretation consistent with the Principle of Relevance will be selected, and the speaker who wants to be correctly understood must make sure that the interpretation she wants to convey is the first one consistent with the Principle of Relevance." Sperber and Wilson do not explain how this might be done with as little effort as seems to be generally expended in informal language use.

[4]Grice suggests (1975:49) "that anyone who cares about the goals that are central to conversation/communication (e.g., giving and receiving information, influencing and being influenced by others) must be expected to have an interest, given suitable circumstances, in participation in talk exchanges that will be profitable only on the assumption that they are conducted in general accordance with the maxims." This seems to me as much a generalization as Sperber and Wilson's Principle of Relevance: "Every act of ostensive communication communicates the presumption of its own optimal relevance" (1986:158).

[5]And all the maxims, unless this absolutely cannot be reconciled with the speaker's behavior.

[6]There is some circumstantial evidence supporting this. First, there are the ubiquitous hedges and the apologies (pp. 47, 49) pounced upon by those who would circumscribe the scope or significance of Grice's programme. Second, there is the fact that Grice (apparently) was himself unsatisfied with the account, and had to be cajoled into letting lectures II and III be published 8-10 years after they were written.

[7]My understanding (or interpretation) has benefited in ways I am sure I cannot trace, from discussion with Jerry Morgan, and from his work and that of Robin Lakoff, Larry Horn, and Penelope Brown and Steve Levinson. No ideas represented here that they would wish to disclaim should be attributed to them.

[8]For a non-technical discussion of this recursion, see Green 1989:13-15, et passim.

[9]Careful reading of "Logic and Conversation" shows that Grice cited quarrels and letter writing as instances that did not fit an interpretation of the CP that he rejected (Grice 1975:48), not one that he supported.

[10]For a discussion of the relevance of the CP and the maxims to feline behavior, see Davison (1990).

[11]These sentences are from Walter Farley's *The Black Stallion Revolts* (New York: Random House, 1953), C.S. Lewis's *The Lion, the Witch, and the Wardrobe* (New York: Macmillan, 1950), and Mary Norton's *The Borrowers* (New York: Harcourt, Brace and World, 1952) respectively. They were chosen to match the excerpt in (6) for syntax, anaphora, and introduction of noun phrases with definite articles.

[12]Cf. Lakoff (1973) and Brown and Levinson (1987:2).

[13]Cf. Green (1988) for an extended discussion of this.

[14]Addressing the UN in Geneva on 13 December 1988, he said, "I condemn terrorism in all its forms, and at the same time salute those sitting before me in this hall who, in the days when they fought to free their countries from the yoke of colonialism, were accused of terrorism by their oppressors."

REFERENCES

Allwood, J., Anderson, L. G. and Dahl, Ö. 1977. Logic in language. Cambridge, England: Cambridge University Press.

Appelt, D. 1985. Planning English sentences. Cambridge, England: Cambridge University Press.

Blum-Kulka, S. and Olshtain, E. 1986. Too many words: Length of utterance and pragmatic failure. Studies in Second Language Acquisition, 8:165-180.

Brown, P. and Levinson, S. 1978. Universals in language usage: Politeness phenomena. Questions and politeness: Strategies in social interaction, ed. by E. Goody, 56-311. Cambridge, England: Cambridge University Press.

Brown, P., and Levinson, S. 1987. Politeness. Cambridge, England: Cambridge University Press.

Brown, G. and Yule, G. 1983. Discourse analysis. Cambridge, England: Cambridge University Press.

Cohen, P. and Levesque, H. 1980. Speech acts and the recognition of shared plans. Proceedings, Canadian Society for Computational Studies of Intelligence.

Cohen, P. and Levesque, M. 1990. Rational interaction as the basis for communication. Intentions in communication, ed. by P. Cohen, J. Morgan, and M. Pollack, 221-256. Cambridge, Mass: MIT Press.

Cohen, P. and Perrault, C.R. 1979. Elements of a plan based theory of speech acts. Cognitive Science, 3:177-212.

Davison, A. 1990. Do cats have propositional attitudes. Manuscript in progress.

Green, G. M. 1982. Linguistics and the pragmatics of language use. Poetics 11:45-76.

Green, G. M. 1988. Some remarks on why there is implicature. Studies in the Linguistic Sciences, 17:2:21-42.

Green, G. M. 1989. Pragmatics and natural language understanding. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Grice, H. P. 1957. Meaning. Philosophical Review, 66:377-388.

Grice, H. P. 1975. Logic and conversation. Syntax and semantics, vol. 3: Speech acts, ed. by P. Cole and J. Morgan, 41-58. New York: Academic Press.

Horn, L. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. Georgetown University Roundtable on Languages and Linguistics, ed. by D. Schiffrin, 11-42. Washington, D.C.: Georgetown University Press.

Lakoff, R. 1973. The logic of politeness: Or minding your P's and Q's. Papers from the Ninth Regional Meeting of the Chicago Linguistic Society, ed. by C. Corum, T. C. Smith-Stark, and A. Weiser, 292-305. Chicago: Chicago Linguistic Society.

Leech, G. 1983. Principles of pragmatics. London, England: Longman.

Lycan, W. G. 1984. Logical form in natural language. Cambridge, Mass.:MIT Press.

Lyons, J. 1977. Semantics. Cambridge, England: Cambridge University Press.

Morgan, J. L. 1978. Two types of convention in indirect speech acts. Syntax and semantics, vol. 9: Pragmatics, ed. by P. Cole, 261-280. New York: Academic Press.

Myers, F. 1984. The language of political elites. Paper presented at the LSA/AAAL winter meeting, Linguistic Society of America/American Association of Applied Linguistics. December, 1984.

Nunberg, G. 1978. The pragmatics of reference. Bloomington, Ind.: Indiana University Linguistics Club.

Nunberg, G. 1988. Poetic and prosaic metaphor. Theoretical issues in natural language processing, 3, New Mexico SU, ed. by Association for Computational Linguistics, 177-180. New York: Association for Computing Machinery.

Perrault, R. C. 1990. An application of default logic to speech act theory. Intentions in communication, ed. by P. Cohen, J. Morgan, and M. Pollack, 161-186. Cambridge, Mass: MIT Press.

Pratt, M. L. 1981. The ideology of speech-act theory. Centrum (N.S.) 1:1:5-18.

Reddy, M. 1979. The conduit metaphor–a case of frame conflict in our language about language. Metaphor and thought, ed. by A. Ortony, 284-324. Cambridge, England: Cambridge University Press.

Ruhl, C. 1989. On monosemy. Albany: SUNY Press.

Schauber, E. and Spolsky, E. 1986. The bounds of interpretation. Stanford: Stanford University Press.

Sperber, D. and Wilson, D. 1986. Relevance: Communication and cognition. Cambridge, Mass.: Harvard University Press.

Stubbs, M. 1983. Discourse analysis. Chicago: Chicago University Press.

Conversational Cooperation in Social Perspective

John J. Gumperz
University of California, Berkeley

The many recent linguistic studies that build on Gricean notions of inference can roughly be grouped into two general categories: (1) work in theoretical pragmatics that concentrates on formal models of inference at the sentence or clause level and (2) empirical sociolinguistic studies that seek to document the workings of specific conversational maxims through studies of speaking rules and practices in various languages, speech communities and discourse genres. My own perspective differs from both of these traditions in that I focus on general and in part universal processes of conversational inference at the level of discourse or, more specifically, at the level of speech exchanges. The basic concern is with the notion of situated understanding, which – in contrast to what one might call 'lexical meaning' – always rests, as I want to argue, on inferences that build on assumptions about conversational cooperation, assumptions that although not formal or readily formalizable are nevertheless implicature-like, contingent assessments of communicative intent.

Although Grice is quite clear about his claim that implicature presupposes assumptions about conversational cooperation, most of the discussion in his seminal article 'Logic and Conversation' (1975, 1989), as well as in his later comments on the topic, focuses on conversational maxims as such and on the inferential logic that underlies meaning assessments. The maxims of quality, quantity, relevance and manner, as well as the notion of implicature, are clearly defined. But no detailed description of what is meant by conversational cooperation is given. Nevertheless, I find the fact that Grice links inferencing to assumptions about conversational cooperation highly suggestive since it can be argued that this move constitutes a major step towards the development of a social perspective on language and understanding. Discourse analysts however have so far hardly begun to explore the significance of this linkage and what it implies for the way we formulate research questions and do our analyses.

Let me begin, then, by elaborating on the notion of cooperation. At the risk of repeating what everyone already knows, I quote from Grice's initial description of what is meant by *implicature*. He refers to what he is attempting to do as an inquiry into 'the general conditions that in one way or another apply to conversation as such, irrespective of its subject matter.' Then, somewhat later, in introducing his discussion of conversational implicatures, he goes on to say:

> Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction. This purpose or direction may be fixed from the start (e.g., by an initial proposal of a question for discussion), or it may evolve during the exchange; it may be fairly definite, or it may be so indefinite as to leave very considerable latitude

to the participants (as in a casual conversation). But at each stage, *some* possible conversational moves would be excluded as conversationally impossible. We might then formulate a rough general principle which participants will be expected (ceteris paribus) to observe, Namely, Make your conversational contribution such as is required at the stage at which it occurs by the accepted purpose or direction of the talk exchange in which you are engaged. (Grice 1989, p.26)

I read the above remarks as arguing for, or at least suggesting, that we take an interactive view of communication where prototypically the source data for analysis is not a single speaker's utterance but a speech exchange or a set of exchanges. What this implies can be summarized as follows. (a) Communicating is not just simply a matter of individuals uttering sentences or speech acts or expressing their thoughts. Speech is produced and evaluated within the context of a communicative ecology involving the cooperative efforts of speakers and respondents. (b) Since communication consists of intentional acts, cooperativeness must be assessed with reference to some commonality of purpose or mutual agreement as to the general direction that an exchange is expected to take. As Grice points out, however, this does not mean that participants must agree in all respects; 'their ultimate aims', as Grice puts it, 'may of course be independent or even in conflict' (1989, p.29). (c) However, the agreement on goals, no matter how general, once achieved, directly affects language usage and, presumably also, the actual inferences made as part of the exchange. It would follow that assessments of a participant's cooperativeness must also be grounded in the organizational properties of the exchange.

But Grice does not specify what these properties might be. His illustrations rely on hypothetical data, and examples, for the most part, consist of single utterances or very short exchanges. Perhaps this is why much of the research on inference in pragmatics has not dealt with longer stretches of discourse. Yet lines such as the following, taken from the discussion of cooperative transactions, clearly refer to longer sequences of exchanges: 'There is some sort of understanding (which may be explicit but which is often tacit) that, other things being equal, the transaction should continue in appropriate style unless both parties are agreeable that it should terminate. You do not just shove off or start doing something else' (1989, p.29). I take it therefore that implicature is also meant to apply at the level of discourse.

What exactly then does cooperation involve? Since conversational maxims occupy such a prominent place in the treatment of inferential processes, one might conclude that cooperativeness can be assessed solely with reference to the maxims. But Grice normally speaks of the cooperative principle *and* the maxims, which suggests that he regards the former as separate from the latter, so that cooperation can perhaps best be understood as collaboration in the pursuit of shared communicative goals which serves as a precondition for maxim-based inferences. But this raises yet another set of questions. Grice's use of terms like quantity, quality, relevance and manner implies that following maxims is for him a matter of choosing appropriate expressions at the level of content. Can cooperation, then, be similarly described in terms of conforming to, violating or disregarding norms of interpersonal relations? The example of a speaker who replies: 'My lips are sealed', in response to a prior speaker's attempt to elicit information suggests that this is perhaps what he proposes. But elsewhere he points out that within the context of

shared assumptions, participants can also have conflicting goals, so that cooperation cannot be purely a matter of content.

I would like to suggest that we try to avoid some of these problems by dealing with shared assumptions or presuppositions at the level of illocutionary force. Cooperation, then, does not just depend on personal intent. Rather, assumptions about cooperation come to be seen as emerging from or inferable from what we may call the organizational properties of the exchange, properties that depend in part on content but in part also on certain features of surface linguistic form. I will illustrate what I mean on the basis of three examples extracted from audiotapes of naturally occurring talk: a cross-examination in court, a broadcast of a political debate, and an intake interview recorded in an employment office.

My approach builds in part on the work of conversational analysts, who during the last two decades have produced many ground-breaking empirical studies to show that speech exchanges everywhere conform to universal principles of conversational management. The basic concern in this work is with Goffman's notion of conversational involvement (1963), which in some ways parallels Gricean conversational cooperation. But active participation on the part of speakers and listeners, as precondition for the maintenance of conversational exchange, is the central concern, and analysis focuses on the strategies by which this is accomplished. Yet, I differ from those conversational analysts who have been interested largely in the general characteristics of exchange sequences in that I argue that involvement depends on inferences that participants must make in order to judge what they hear as cohesive talk, and that these inferences in turn are significantly affected by perception and interpretation of certain types of linguistic signs or contextualization cues (Gumperz 1982, 1990).

Prototypically, the initial unit in my analysis of contextualization processes is an *event*, defined as a temporally ordered sequence of exchanges characterized by a detectable beginning and an end which provides empirical evidence of what the event's outcome is, and therefore also evidence to confirm or disconfirm the analyst's assumptions about what was intended. My claim is that understanding in such events rests on two levels of inferencing: a global level where what is at issue is the overall purpose or direction of the exchange and a local level of inferencing which concentrates on sequential relationships such as what the conversational analysts call 'adjacency pairs'. Here I will illustrate the approach in the course of the actual analyses. For a detailed explanation, see Gumperz 1982, 1989.

The first example comes from cross-examination testimony by a rape victim (Drew 1990). Here is the relevant background. The victim had been sitting at a table with some friends at a local club one evening. The accused, who had come to the club alone, had joined the group and participated in the general conversation. When the victim got up to leave, the accused walked out with her, offered her a ride and asked her to accompany him for a cup of coffee. A few minutes after leaving the parking lot, he had turned into a side road and driven to an isolated spot where the attack occurred. In the following transcript, C is the defense counsel cross-examining W, who is the victim. (For explanation of transcription conventions used below, see Appendix.)

I. Rape Trial Cross-Examination.

1 C: well you had some uh p- .. uh fairly lengthy conversation
2   with the defendant, uh didn't you?
3   <0.7>
4 C: on that evening of February fourteenth?
5   <1.0>
6 W: we:ll we were *all talking.
7   <0.8>
8 C: well *you kne:w at that ti:me that the defendant was
9   *in:terested .. in *you, .. didn't you?
10   <1.3>
11 W: he: asked me how I'(d) *been:, and
12   <1.1>
13 W: j- just stuff like that

Some turns omitted.

14 C: and you went to *a: uh <0.9> ah you went to a *ba:r?
15   in *Boston <0.6> is that correct?
16 W: it's a *clu:b/
17 C: it's where uh .. uh <0.3> gi:rls and fella:s *meet, isn't it?
18 W: people *go there/
19 C: and during that eve:*ning:
20   didn't mister O come over to sit with you?
21   <0.8>
22 W: sat at our *table/

Some turns omitted.

23 C: some distance back into the uh .. *wood, wasn't it?
24   <0.5>
25 W: it was up the *path/ I don't know how far/

Anyone examining the preceding encounter one exchange at a time might conclude that conversational cooperation has broken down since the witness regularly fails to provide direct answers to simple yes/no questions, disputes the counsel's use of words and in general seems to be interfering with his efforts to produce a coherent account. But more detailed examination suggests that the situation is more complex. We know that the underlying goal of courtroom interrogation is to establish the facts of the case for the benefit of the judge and jury. Moreover, in contrast to direct examinations where questioning is, for the most part, supportive, cross-examinations are adversarial proceedings where counsel's questions are designed to uncover flaws or inconsistencies in the witnesses' testimony, while the witnesses in turn seek to establish or reaffirm their own version of what took place. The two participants, therefore, can be seen to be engaged in a contest over which of several possible interpretations should be accepted as the correct one. Furthermore, given the rules of courtroom procedure, where turns at speaking are strictly allocated and participants are highly constrained in what they can or cannot put into words, the contest must in large part rely on indirectness. That is to say, what the dispute over interpretation is about must be inferred indirectly from the two speakers' language use, in particular, their choice among sets of related expressions.

An examination of the surface content of the exchange brings out the dynamics of the contest. The counsel's questions in lines 1-3 and 4 can be understood as suggesting the implicature that the witness and the defendant knew each other well before the incident, so that perhaps there might have been a prior sexual relationship between them. The witness is, therefore, faced with a dilemma. To answer 'yes' would be to confirm the counsel's implied meaning, but to say 'no' could likely have been shown to be untruthful. By substituting her own expression for counsel's wording, she manages to neutralize the implicature. Her answers to the following two questions in lines 11 and 13 have a similar effect. In lines 15 and 18 counsel seeks to depict the witness as a woman who frequents questionable places to look for men and could have been looking for sexual adventures. Again she counters by substituting the more neutral 'club' for 'bar' and 'people go there' for the sexually suggestive 'girls and fellas meet'.

Yet, how do we explain the fact that interpretation differs depending on whether we examine the exchanges one at a time or whether we look at the interaction as a whole? The answer must lie in the changing nature of the contextual presuppositions that enter into the interpretive process. Had the witness disputed the counsel's wording in just one exchange, her behavior might have simply been regarded as slightly odd, or perhaps nitpicking or somewhat argumentative. It is the cumulative effect of her responses that leads listeners to question their initial interpretation and draw on their knowledge of what a cross-examination is about and of the conflicting aims of the two participants to construct a new envisionment in which to ground their understanding of what the facts are. In other words, what we are talking about here are schema-based inferences where the implicatures that the counsel is trying to depict the witness as someone who willingly consented to the act and that the witness seeks to present herself as an innocent victim of an unprovoked attack rest on the presupposition that, to paraphrase Grice, each participant recognizes 'a common purpose or set of purposes' or 'a mutually accepted direction' in the exchanges. Both individuals agree that the goal is to convey to the jury what the facts of the case are, although each of the two is indirectly suggesting that the other's facts are wrong.

It is important to note that each participant's action at any one point is dependent upon that of the other. It is only in this way that we can understand responses like 'He asked me how I'd been – just stuff like that,' in answer to 'You knew that the defendant was interested in you,' and 'Sat at our table,' in answer to 'Didn't Mr. O come over to sit with you?' as implicating that the witness and the defendant only knew each other casually. This interdependence of participants' actions is not always apparent from surface form. Since the rules of courtroom discourse limit what can and cannot be lexicalized, speakers rely on non-verbal cues to convey inter-sentential relationships.

A brief look at both speakers' use of prosody and of syntactic form illustrates this point. For example, 'bar' and 'club' in lines 14 and 16 both carry similar stress and fill equivalent syntactic slots. This is also the case for the verbs 'meet' and 'go' in lines 17 and 18, and the nouns 'wood' and 'path' in lines 23 and 25. This is to say, then, that the listener perceives that the two terms are being contrasted from the way in which the talk is contextualized. The contrast, then, in turn suggests the

inference that the witness is intentionally substituting one term for the other, a fact on which the implicature rests.

It is also interesting to examine how pausing strategies are employed in light of the two contestants' different goals within each one's understanding of the overall purpose of the interaction. Note that there is a .7 second pause after the counselor's initial inquiry which he then follows with a clarification question. This in turn is followed by a one second pause before the witness answers. Similarly, her remarks in line 11 and 13 are preceded by long pauses. If we take the witness's perspective, such pausing can be interpreted as as an attempt to gain time to think so as to avoid possible traps on the part of the counselor. But the counselor might well claim that the witness's pausing indicates that she is procrastinating or is unsure of her answers. And it is possible to argue that in line 4 he cuts her short in order to convey the impression to the jury that she is in fact procrastinating. The counselor's own pausing, on the other hand, is not subject to such ambiguities. All of this again follows from the presupposition that the two contestants are engaged in a cross-examination.

To sum up then, it is clear that the two speakers understand each other and are cooperating when cooperation is understood at the level of illocutionary force. Cooperation, then, is both willingness and ability to collaborate in the production of coherent discourse. In the rest of this paper, I would like to illustrate some of the implications of this perspective.

Example II comes from a transcript of a 1960's radio broadcast of a panel discussion involving the then mayor of San Francisco, Joseph Alioto (A), and two leaders of the local black community, B, who is not heard in this excerpt, and Eldridge Cleaver (C). The mayor has been seeking to convey a favorable picture of what his administration has been doing in encouraging the black community to participate in municipal decision-making. B objects to this characterization. The excerpt begins as the mayor is responding to him.

II. Panel Discussion: March 7, 1968.

```
1 A:  but the *point **i:s,
2     .. that the *black community in San Francisco Hunter's Point/
3     is *not giving me the story that they're being excluded
4     from any kind of positions of responsibility/
5     .. and i *have *told them,
6     if they're gonna par*ticipate in these things/
7     that there're gonna be *no,
8     .. bequeaths from Mt Olympus down *to them//
```

B's relatively brief turn is not included here. The transcript picks up as C breaks in, taking over B's turn with a latched response in which he directly contests the mayor's claim, arguing that, far from trying to involve the black community in local decision-making, the mayor's administration, along with other municipal administrations throughout the country, has been engaged in a war against the black community.

 9 C: =={[lo] i would just like to say *this/}
10   that uh .. the mayor, ... is the mayor of San Fran*cisco/
11   he's not the mayor of the United *States//
12   .. and uh, that's his area of responsi*bility//
13   ==an::d, i would like to take exception to uh,
14   ... the glowing picture that he's just painted,
15   .. as to the relationship between uh,
16   .. his *office, and his admini*stration,
17   and his *government, .. and the black community//
18   now it may well *be that he has,
19   a good rapport with certain people in Hunter's Point//
20   .. {[hi] at the same *ti:me,}
21   .. the black community is very well a*ware/
22   ... that San Francisco, .. *under Mayor A-
23   ... {[ac] A, excuse me/}
24 A: {[pp] alright//}
25 C: uh ... the same as a::ll,
26   .. uh city governments across this *country/
27   ... {[dc] are sharply escalating/}
28   ... the war against the black community//
29   that the po*lice, ... who occupy black communities/
30   .. are becoming {[dc] *more and *more}
31   of an occupying army, {[ac] *in the black communities/}

At this point, A breaks into C's long turn with a latched challenge. There fol-
low two brief exchange sequences until C resumes the presentation of his position.

32 A: ==in San Francisco?
33 C: =={[f] in San Fran=cisco//}=
34 A: =aw th=at's just a generalization/ ={[lo] that's not true//}=
35 C: ={[hi] in San Fran*cisco/}=
36 A: ={[lo] that just isn't true//}=
37 C: =in San Fran*cisco/= the black population, .. is well aware,
38   a:nd they're very very much con*cerned,
39   .. uh with the type of tactics/
40   a:nd the increasing {[ac] bru*tality of the San Francisco police department/}
41   a:nd the *fact that the San Francisco police department/
42   goes around *kicking down people's *doors/
43   .. entering, .. il*legally without search warrants/
44   .. with*out warrants for arrest//
45   uh you may say that/
46   .. uh the black community is cooperating with you fully/
47   and that they're not uh, con*cerned about this/
48   .. but i know for a *fact/ that the =black community=

Here again A breaks in with a question which seeks to categorize what C has
just said as a 'complaint.' That is, by implication, a typical instance of the many
problems that administrators have to deal with as part of doing their job.

49 A: =they have complaints?= is that what you mean?

50 C: complaints?
51 A: =they have complaints, the *white community has com*plaints,=
52 C: =they're not complaints/ they're not com*plaints//=

When C questions A's use of the term 'complaints' in line 49 by echoing the latter's words with a rising intonation, A repeats what he had said in line 49 with a slightly falling intonation on 'complaints' such as is found in listing, and then begins to list the many complainers that he has to deal with. C then tries to shift the rhythmic patterns with his repeated statements in line 52 so as to deny the applicability of the term 'complaints' to the black community's situation. His turn overlaps A's previous turn, but A continues with his rhythmic string of parallel constructions involving repeated mention of the theme of complaints, which leaves no space for C's counter-arguments.

53 A: the *North Beach community has com*plaints,
54   the *straight colony out in Haight-Ashbury has com*plaints,
55   the fact that there are com*plaints from different communities,
56   .. doesn't mean that we have some kind of
57   irreconcilable conflict that **must erupt in violence//
58   ={[hi] it doesn't mean that at all//}=

As in example I, the two speakers agree on the general goals or purposes of the exchange. They know that to debate is to engage in a contest over which of several conflicting views will prevail. What distinguishes this encounter from example I, however, is the way in which contextualization strategies enter into the argumentation. At the outset A employs accent and choice of terminal intonational contours to achieve rhetorical effect. C on the other hand relies on phrasing, pausing and pitch register shifts to chunk the stream of speech in such a way as to highlight his points. In line 37, C manages to abort A's challenge by means of the rhythmicity of his delivery. But towards the end of the passage, A shifts his tactics and adopts C's way of using repetition and rhythm as a contextualization strategy. By doing so, he manages to drown out C's objections. In other words, C automatically responds to A's new contextualization strategy and allows A to go on with his argument. Basically, each speaker is aware of what the other is trying to do on the level of intent, but A alone seems to have control over both strategies at the level of discourse production and this gives him the advantage in the debate.

The next example consists of a passage taken from an initial interview at a state-financed employment center where job seekers register for positions. The encounter was recorded in the British Midlands. The interviewer (A) is a native speaker of the local form of British English, and the candidate (B) a locally resident native speaker of a South Asian language with relatively good instrumental control of English, whose discourse strategies, however, reflect those of the native South Asian language. The transcript begins shortly after the introductory portion of the encounter as the interactants turn to systematic questioning.

III. Job Center Interview.

8   A: yeah, you're on the youth opportunity .. scheme (xxx)?
9   B: .. yeah/
10 A: (xx) that's right/ can you tell me a little bit about it?
11   what you what you're actually doing (on it)?

12 B: well ehm .. you do .. two weeks,
13 A: .. hmm/
14 B: we change around every two weeks/
15 A: yeah/
16 B: we're doing ehm .. plastering now/
17 A: .. yeah/
18 B: it's a work (   ) opportunity/
19 A: oh yeah, yeah/
20 B: and then ehm ... we've done some plumbing,
21 A: yeah/
22 B: and next week, we'll be going to eh painting, ... wallpapering,
23 A: .. oh yeah/
24 B: (xxxx)/
25 A: yeah/
26 B: (xxx)
27 A: alright now, just going back to what you said at first, the plastering,
29 B: (   )
30 A: how- .. you know, what- what do you actually do?
31   how- how involved are you?
32 B: well, ... he shows us what to do,
33   and we ... plaster the wall out in plasterboard,
34 A: yeah/ do you do all the mixing on that?
35 B: ==yeah/
36 A: ==you do all that?
37 B: we mix it by hand,
38 A: ... mhm/
39 B: and then we-
40   ... we (   ) wall ( ) to learn how to plaster/
41 A: yeah/ right/ and it is- it- .. are you just putting the- .. y'know,
42   skimmin, ... and that? i mean, are are you doing are you doing
43   everything?
44 B: yeah/
45 A: you're doing it all/ yeah/ ... do you do any- .. working at
46   heights? anything like that?
47 B: yeah, we use scaffold/
48 A: ... yeah/
49 B: ==and then, ... we work on scaffold/
50 A: yeah, what sorts of heights have you worked in?.. (   ) like?
51 B: like .. fifteen foot,
52 A: ==say fifteen/
53 B: ==something like that/ .. (   )
54 A: ==ri:ght/ and .. you
55   said, .. before that, you'd done .. plumbing, did you say?
56 B: yeah/
57 A: can you tell me a little bit about what you've done.. on that
58   side?
59 B: well, we did it at nelson/

At first glance, the above exchange reveals little that strikes one as problematic. The content shows that both interviewer and applicant are aware that what is at issue at this point is the gathering of information on the applicant's previous employment and training history. Although the interviewer at times has problems in eliciting the information she needs, the applicant's answers seem to be at least minimally informative and thus cooperative.

However, if we look in detail at the sequencing of turns at speaking, some significant problems become apparent. In line 10, A asks for information on what the applicant is 'actually doing on the youth opportunity scheme', a type of question that is commonly understood as a request to list and describe the job skills the applicant has acquired. The answer 'You do two weeks', however, makes no reference to skills. A then pauses as if to wait for B to elaborate before interjecting 'hm', which, given the context, counts as a mild prompt. But B simply responds by rephrasing his earlier statement. A responds with 'Yeah'. This then elicits 'We are doing plastering now.' Although plastering is mentioned here, the syntactic form of the utterance shows that what is being referred to is a general activity, not a specific skill. There follows another pause followed by the interviewer's 'Yeah'. When B then responds 'It's a work opportunity,' A's reply, 'Oh yeah, yeah', suggests impatience. In his next turn then, B refers to yet another activity, plumbing, and then responds to the next 'Yeah' by mentioning a third activity, wallpapering. In reply A once more pauses to give B another chance to produce more relevant detail. When there is no result, her 'Oh yeah' indicates increasing impatience. In response B lowers his voice and his talk is mumbled and unintelligible. When A repeats her prompt, B produces another unintelligible answer.

Finally, A attempts once more to get a more responsive reply, this time making explicit what she wants in words, 'Going back to what you'd said at first, the plastering, what do you actually do? How involved are you?' B responds by being just a little bit more specific about the techniques they learn in the course. But A continues to have difficulty in eliciting the specific data she needs. No information is spontaneously given. A has to work for every bit of relevant information. Her question in line 41 reveals what kind of replies she is looking for: statements that make specific reference to the skills that make up the various operations that are mentioned, and that refer to these components using conventional terminology employed by the relevant craftsmen, such as 'skimming', 'mixing plaster', 'working on scaffolds', 'plasterboarding', etc., and presumably listed in job specifications.

What is at issue here is that the applicant's responses violate the maxim of relevance, with 'relevance' being defined here in terms of what we know about the communicative expectations associated with this type of job interview. As with the lexical substitutions in example I, it is the cumulative effect of such violations that suggests that lack of cooperativeness. Again, one single instance might have passed as a lapse. In contrast to example I, however, the two interactants do not have conflicting goals. Moreover, the applicant is at first unaware of the fact that his answers are problematic. When, as with A's exclamation in turn 23, it becomes evident that B is not giving the expected answers, he seems distressed, as evidenced by his mumbled replies.

One problem with the B's responses is that he is not aware that the interviewer needs specific information about the actual skills that he has acquired in order to

determine what job openings he may be eligible for. But even if he did know this, since he does not seem to control the technical vocabulary in terms of which to describe what he can do, his answers are found lacking. In addition, he seems to be unable to apply his knowledge of what the purposes or goals of the encounter are to his interpretation of what is transpiring at the level of utterance sequences. As a result, he fails to recognize the interviewer's hints and prompts for what they are.

The interactants have evidently failed to achieve conversational cooperation. But the issue is not one of unwillingness to cooperate or differences in agenda at the level of activity. The problem lies in signalling at the level of speech act sequential organization. The two participants seem to be operating with different sets of contextualization conventions, so that the interviewer's prompts and her attempts at cuing the applicant with relevant terms are misunderstood.

The analysis shows that if we take the notion of speech exchange seriously and adopt an interactive perspective along the lines of that outlined at the beginning of this paper, then conversational cooperation becomes considerably more complex than would appear from a surface reading of Grice's 'Make your conversational contribution such as is required at the stage at which it occurs by the accepted purpose or direction of the talk exchange in which you are engaged' (1989, p.26). Cooperating, in other words, requires much more than simply agreeing to engage in a collaborative endeavor. To begin with, while participants as individuals may have shared understandings of what they seek to achieve, these understandings must be realized by engaging in one or another activity type (Levinson 1979). Activity types are associated with culturally defined, communicative goals, knowledge of which is acquired in the course of an individual's communicative experiences. Cross-examinations, political debates, and job interviews are instances of such activity types. Such context-specific, shared understanding of what the activity's goals are, along with knowledge of relevant contextualization cues, other linguistic knowledge, and other matters then enter into implicatures that underlie interpretations of specific exchanges. In example I, it is the rape case witness's skillful use of lexical substitutions to generate such implicatures that makes her strategies so successful. In example II, what is significant is the interactants' use of contextualization strategies, particularly of prosody and rhythm. These contextualization strategies constitute embodied, taken-for-granted knowledge that is automatically applied and responded to without conscious reflection. When C, the second debater, automatically responds to his opponent's use of such strategies by yielding the floor, he loses control of the debate. In example III, cooperation breaks down in spite of the fact that both parties share a common purpose, and both intend to cooperate. It is apparent that, despite his willingness to cooperate, the applicant is unable to do so because he lacks relevant background knowledge and does not have sufficient control of contextualization strategies.

The notion of cooperation that emerges in this differs significantly from what Grice intended. One might, for example, argue that Goffman's term *coparticipation* is more applicable to the phenomena I describe. But clearly the inferential processes referred to here are implicature-like. Once we take an interactive view of speech exchanges, we also leave the realm of individual intent and are forced to deal with shared, mutually understood knowledge. Cooperation, then, cannot be described in commonsense terms. It becomes a precondition for the production of

coherent discourse which crucially involves both willingness to collaborate, as a matter of personal intent, and ability to collaborate, as a function of communicative competence.

Appendix

Symbol	Significance
//	Final fall
/	Slight fall indicating "more is to come"
?	Final rise
,	Slight rise as in listing intonation
-	Truncation (e.g. what ti- what time is it/)

..	Pauses of less than .5 second
...	Pauses greater than .5 second (unless precisely timed)
<2>	Precise units of time (= 2 second pause)
=	To indicate overlap and latching of speakers' utterances

     e.g. R: so you understand =the requirements=

           B: =yeah, i under=stand them/

       ~~~~~~~~~~~~~~~~~~~~~~~~~~

           R: so you understand the requirements?

           B: ==yeah, i understand them/

           R: ==and the schedule?

           B: yeah/

with spacing and single "=" before and after the appropriate portions of the text indicating overlap and with turn-initial double "=" indicating latching of the utterance to the preceding one.

::	Lengthened segments (e.g. wha::t)
*	Accent; normal prominence
**	Extra prominence
{[ ]}	Nonlexical phenomena, both vocal and nonvocal, which overlay the lexical stretch; [ac] accelerated; [dc] decelerated; [lo] low pitch register; [hi] high pitch register; [p] quiet voice; [pp] very quiet voice; [f] loud voice; [ff] very loud voice; e.g. {[lo] text//}

( )	Unintelligible speech
di(d)	A good guess at an unclear segment
(did)	A good guess at an unclear word
(xxx)	Unclear word for which a good guess can be made as to how many syllables were uttered with "x" = one syllable

441

References

Atkinson, J. M. and Heritage, J. (eds.). 1986. Structures of Social Action. Cambridge: Cambridge University Press.

Drew, P. 1990. Disputes in Court Room Cross-examination: 'Contrasting Versions' in a Rape Trial. Rethinking Context, ed. by A. Duranti and C. Goodwin. New York: Cambridge University Press.

Goffman, E. 1963. Behavior in Public Places: Notes on the Social Organization of Gatherings. Glencoe, IL: The Free Press.

Goodwin, C. 1986. Notes on Story Structure and the Nature of Participation. Structures of Social Action, ed. by J. M. Atkinson and J. Heritage. Cambridge: Cambridge University Press.

Goodwin, C. and Goodwin, M. H. 1991. Frameworks for Participation Within Units of Talk. Contextualization, ed. by P. Auer and A. Di Luzio. Amsterdam: John Benjamins. In press.

Grice, H. P. 1975. Logic and Conversation. Syntax and Semantics, vol. 3: Speech Acts, ed. by P. Cole and J. Morgan, 41-58. New York: Academic Press.

Grice, H. P. 1989. Studies in the Way of Words. Cambridge, MA: Harvard University Press.

Gumperz, J. J. 1982. Discourse Strategies. New York: Cambridge University Press.

Gumperz, J. J. 1989. Contextualization and Understanding. Berkeley Cognitive Science Report No. 59, April 1989.

Gumperz, J. J. and Berenz, N. 1990. Transcribing Conversational Exchanges. Transcribing and Coding Methods for Language Research, ed. by J. A. Edwards and M. D. Lampert. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. In press.

Heath, C. 1986. Body Movement and Speech in Medical Interaction. Cambridge: Cambridge University Press.

Levinson, S. C. 1979. Activity Types and Language. Linguistics, vol. 17:356-399.

Givenness, Implicature, and the Form of Referring Expressions in Discourse

Jeanette K. Gundel
University of Minnesota

Nancy Hedberg
Cornell University
Simon Fraser University

Ron Zacharski
University of Minnesota

Grice's Cooperative Principle and the associated maxims attempt to provide an explanation for how it is that we can communicate more than we literally say. In this paper, we outline a theory that builds on Grice's work as a basis for explaining the choice of coding for referents in natural language discourse.

It is generally recognized that the form of natural language expressions correlates with cognitive statuses such as givenness, topic and presupposition, which a cooperative speaker can assume on the part of the addressee. In two previous papers (Gundel, Hedberg and Zacharski 1988, 1989), we proposed that there are five cognitive statuses relevant to the form of referring expressions in English discourse, and that these are related in the Givenness Hierarchy shown in (1). In the present paper, we extend the investigation to six additional languages - Arabic, Chinese, Japanese, Korean, Russian, and Spanish.

(1)   **The givenness hierarchy**[1]

in focus	>	activated	>	familiar	>	uniquely identifiable	>	type identifiable
{*it*}		$\left\{\begin{array}{l}that\\this\\this\ \text{N}\end{array}\right\}$		{*that* N}		{*the* N}		$\left\{\begin{array}{l}a\ \text{N}\\\text{indefinite } this\ \text{N}\end{array}\right\}$

1.   The Givenness Hierarchy

Each status on the hierarchy is a necessary and sufficient condition for the appropriate use of the English form or forms listed under it. Use of these forms thus conventionally implicates that the associated cognitive status is met and, since each status entails all lower statuses, it also conventionally implicates that all lower statuses (i.e. statuses to the right) have been met. Thus, the definite article *the* conventionally implicates 'you can identify this', the demonstrative determiner *that* conventionally implicates 'you are familiar with this', and therefore can identify it, and so on. The statuses are briefly characterized below.

type identifiable   In the weakest case, a form is appropriate only if the speaker can justifiably assume that the addressee is able to identify the type of entity referred to, i.e. the class of objects described by that expression. This status, which we refer to as 'type identifiable', is necessary for appropriate use of any referring expression and it is sufficient for indefinite reference. Thus, *a dog* or *this dog* in (2) is appropriate if the addressee can be assumed to know the meaning of the word *dog*.

(2)   I couldn't sleep last night. $\left\{\begin{array}{l}\text{A}\\\text{This}\end{array}\right\}$ dog next door kept me awake.

uniquely identifiable   Definite descriptions differ from indefinite ones in that their appropriate use requires that the addressee can not only identify the **type** of entity being referred to, but can also associate a representation with the particular entity or entities that the speaker has in mind. This status, which we refer to as 'uniquely identifiable', is a necessary condition for all definite reference, and it is both necessary and sufficient for appropriate use of the definite article *the*. Identifiability may be based on previous familiarity with the referent, but, as Hawkins (1978) and others have pointed out, it doesn't have to be. The basis for identification may be fully encoded in the form itself. Thus, the phrase *the dog next door* in (3) is perfectly felicitous even if the addressee has no previous knowledge that the speaker's neighbor has a dog.

      (3)   I couldn't sleep last night. **The dog next door** kept me awake.

familiar   Although identifiability doesn't require previous knowledge of the referent, it is usually the case that the referent of an identifiable noun phrase is identifiable because both speaker and addressee are already familiar with it, and thus have some mental representation of it. This status, which we refer to as 'familiar', is a necessary condition for use of all definite demonstratives and pronouns, and it is sufficient for the demonstrative determiner *that*. Thus, (4) unlike (3) is felicitous only if the addressee already knows that the speaker's neighbor has a dog.

      (4)   I couldn't sleep last night. **That dog next door** kept me awake.

activated   The set of familiar entities includes those which the speech participants are currently aware of, i.e. have access to, due to their presence in the immediate discourse context. We refer to this status as 'activated'. Activated entities may have been linguistically introduced, or they may be activated by virtue of their presence in the extralinguistic context. Activated entities, therefore, always include the speech participants themselves.

      Activation is necessary for all pronominal forms, and it is sufficient for appropriate use of the demonstrative pronoun *that* as well as for stressed personal pronouns. The pronoun *that* in (5) can thus be used appropriately to refer to the barking of a dog only if a dog has actually been barking during the speech event or if barking had been introduced in the immediate linguistic context.

      (5)   I couldn't sleep last night. **That** kept me awake.

Activation is also necessary for the demonstrative determiner *this*.

      Both determiner and pronominal *this* have the additional condition that the referent be not only activated, but speaker-activated, by virtue of its inclusion in the speaker's context space (cf. Fillmore 1975, 1982, Lakoff 1974). For example, the phrase *this dog* in (6) is inappropriate in the context of A's question.

    (6) A: Have you seen the neighbor's dog?
          B: Yes, and **that dog** kept me awake all night.
          B': ?? Yes, and **this dog** kept me awake all night.

      But in (7), where the dog has been introduced by the speaker, either *this dog* or *that dog* is appropriate.

(7)   My neighbor has a dog. $\left\{ \begin{array}{l} \textbf{This} \\ \textbf{That} \end{array} \right\}$ dog kept me awake.

in focus   Finally the most highly activated entities are not only in the speaker's and hearer's awareness but are also at the center of attention at the current point in the discourse. This status, which we refer to as 'in focus', includes activated entities which are likely to be continued as topics of subsequent utterances. The status 'in focus' thus not only involves the assumed status of an entity in memory but also its relative importance at a given point in the discourse. In focus is a necessary condition for appropriate use of zero and unstressed pronominals.

Thus, since the Bull Mastiff is the topic of (8a), it can be appropriately referred to with either *that* or *it* in (8b). But in (9), where the Bull Mastiff has been introduced in a phrase that functions primarily to restrict the referent of the direct object, reference with the unstressed pronoun *it* is inappropriate.[2]

(8) a.   My neighbor's Bull Mastiff bit George.
    b. $\left\{ \begin{array}{l} \textbf{It's} \\ \textbf{That's} \end{array} \right\}$ the same dog that bit Mary Ben.

(9) a.   Sears delivered new siding to my neighbors with the Bull Mastiff
    b.   **That's** the same dog that bit Mary Ben.
    b'. #**It's** the same dog that bit Mary Ben.
    c.   Anyway, this siding is real hideous and...

2.   The Givenness Hierarchy and Grice's Maxim of Quantity

Since each of the cognitive statuses in the Givenness Hierarchy entails all lower statuses, reference with a particular form also implies the possibility of reference with forms associated with lower statuses. In our earlier examination of definite referring expressions in English data drawn from a variety of spoken and written genres, we found, as predicted, that forms are distributed across all statuses which meet necessary conditions for their appropriate use. For example, there were some tokens of *the N* for all statuses to the left of uniquely identifiable and some tokens of demonstrative pronoun *that* for both activated and in focus.

However, the distribution across statuses varies considerably for different forms, and in many contexts a given form is inappropriate, or conveys some special effect, even when sufficient conditions for its use have been met. We argued that such facts follow naturally from interaction of the givenness hierarchy with Grice's maxim of quantity, stated in (10).

(10)   **Maxim of Quantity** (Grice 1975)
       Q1 Make your contribution as informative as required (for the current purposes of the exchange).
       Q2 Do not make your contribution more informative than is required.

Thus, as Grice himself noted, an indefinite article, which requires only that the referent be type identifiable, conversationally implicates by Q1 (make your contribution as informative as required) that the referent is not uniquely identifiable and hence also not familiar to the addressee. Similarly, even though all referents in focus are also activated, use of a demonstrative pronoun, which requires only activation, generally implicates by Q1 that the referent is not in focus. This

accounts for the frequently noted function of demonstrative pronominals to signal a shift in focus.

On the other hand, use of a particular form doesn't always conversationally implicate that necessary conditions for a form requiring a higher status don't obtain. Over half the tokens of definite article *the* in the English data we examined occur with phrases whose referents are not only identifiable, but also familiar and even activated or in focus, and the overwhelming majority of noun phrases whose referents are familiar but not activated are referenced with *the* N rather than *that* N. Thus, a definite article clearly doesn't implicate that the referent is not familiar. We have argued that this is due to the fact that for full definite NPs, signalling identifiability is sufficient for associating a representation of the referent and an explicit signal of a more restrictive cognitive status is therefore unnecessary. In other words, it is Q2 (don't give more information than required) which applies here.[3] This is especially true when there is a coreferring phrase in the immediate discourse context which is at least partially identical in form, as in (11).

(11)   "How in the world," demanded Harriet, "did *you* get here?"
        "Car," said Lord Peter, briefly. "Have they produced the body?"
        "Who told you about **the body**?"

When the demonstrative determiner *that* is used, it often facilitates comprehension by serving as an explicit signal to the addressee to search long-term memory for a a familiar referent. In such cases, which we have referred to as 'reminder *that*'s, determiner *that* signals (i.e. conventionally implicates) that the referent is familiar and conversationally implicates by Q1 that the referent is not activated. An example is given in (12).

(12)   Exxon oil claims it will take several million dollars to clean up **that oil spill off the coast of Alaska**. [beginning of radio newscast]

Other special effects associated with demonstratives, such as emotional uses discussed, for example, in Lakoff 1974, may also be attributed to quantity implicatures.

3.   Universality of the Givenness Hierarchy

The research described above was based entirely on English, and it is not clear to what extent it can be generalized to other languages. The purpose of our current research is to extend the investigation to six additional languages: Arabic, Russian, Japanese, Mandarin Chinese, and Spanish. The following questions are addressed:

1.   Are the five statuses on the givenness hierarchy necessary and sufficient for describing conditions on the appropriate use of different types of referring expressions in all languages?

2.   Are necessary conditions for the appropriate use of analogous forms the same across languages, or do these differ depending on what forms are available in the languages in question?

3.   When necessary conditions for the use of more than one form are met, does the Givenness Hierarchy interact with the Maxim of Quantity in the same way across languages?

Our data for Russian comes from published transcripts of casual conversation. For the other six languages (including English) we have narrative film descriptions which were collected for another study.[4] These data were supplemented by native speaker judgments and from whatever relevant information we were able to collect from grammars and scholarly articles. Our findings regarding necessary and sufficient conditions for the appropriate use of different forms in these languages are given in the chart in (13) (capitals indicate a stressed pronoun).

(13)

	in focus	activated	familiar	unique ident.	type ident
Arabic	ø, *hua*	*HUA, haadaa, daalika, haadaa* N	*daalika* N	*al* N	ø N
Chinese	ø, *ta*	*TA, zhè, nèi, zhè* N		*nèi* N	ø N (*yi* N)
English	ø, *it*	*HE, this, that, this* N	*that* N	*the* N	*a* N
Japanese	ø	*kare, kore, sore, are, kono* N, *sono* N	*ano* N		ø N
Korean	ø	*i, ku, cə,i* N	*ku* N, *cə* N		ø N
Russian	ø, *on*	*ON, ètot, tot*	*ètot* N, *tot* N		ø N
Spanish	ø, *él*	*ÉL, éste, ése, aquél, este* N	*ese* N, *aquel* N	*el* N	*un* N

Demonstrative Forms

	proximal	medial	distal
Arabic	*haadaa*	*daalika*	
Chinese	*zhè*	*nèi*	
English	*this*	*that*	
Japanese	*kore, kono* N	*sore, sono* N	*are, ano* N
Korean	*i*	*ku*	*cə*
Russian	*ètot*	*tot*	
Spanish	*éste, este* N	*ése, ese* N	*aquél, aquel* N

As seen in the chart, the cognitive statuses posited for English appear to be sufficient for describing conditions on the appropriate use of demonstratives, articles and pronouns in the other six languages we examined. However, not all five statuses are necessary for Chinese, Japanese, Korean and Russian, the languages which lack distinct forms for articles. As illustrated in (14)-(17), a noun with no preceding determiner in these languages can be interpreted as either uniquely identifiable (i.e. definite) or merely type identifiable (ie. indefinite).

(14)   *Ta zai **bisai** zhong huojiang*
           he in   game during win-prize
           "He won a prize in **a/the game.**

(15)   *kare*   *wa*   *akai*   *kingyo*   *o*   *hoshii*
         he   TM   red   goldfish   OM   want
         "He wants **a/the red goldfish**."

(16)   *Ku-nun* *pul-un*   *kumpungo-lul*   *won-ha-n-ta*
         he-TM   red-REL goldfish-OM   want-PRE-D
         "He wants **a/the red goldfish**."

(17)   *V* *ruke*   *derzhali*   *bilety*
         in   hand   held 3pl   tickets
         "In their hands, (they) held **tickets/the tickets**"

These languages differ, however, with respect to which status is unnecessary. As seen in the chart in (13), Japanese, Korean and Russian have no form for which the status uniquely identifiable is both necessary and sufficient. In Chinese, on the other hand, familiarity appears to be both necessary and sufficient for appropriate use of the distal demonstrative determiner *nei*; but Chinese apparently has no determiner which requires that the referent be familiar, but not necessarily activated.

Thus, according to the Chinese speakers we consulted, example (18), unlike its counterpart in the other languages, is appropriate even if the addressee has no previous knowledge that the speaker's neighbor has a dog.

(18)   *tsuotian*   *wanshang*   *w o*   *shui-bu-zhao*
         yesterday   evening   I   sleep-not-
         gebi-de   *n è i*   *tiao*   *g o u* *jiao de*   *lihai*
         next.door   that   clf   dog   bark prt   extremely
         "I couldn't sleep last night. **The (lit "that") dog** next door was barking."

This supports the observation, made for example in Li and Thompson 1981, that the unstressed distal demonstrative in Chinese is beginning to function like a definite article.

We turn now to our second question - are necessary conditions for use of corresponding forms the same across languages? The answer here appears to be yes, for all forms except demonstrative determiners.[5] Thus, in all these languages, unstressed personal pronouns and zero anaphora require that the referent be in focus; demonstrative pronouns and stressed personal pronouns require only activation; the definite article requires only unique identifiability; and the indefinite article requires only type identifiability.

For demonstrative determiners, on the other hand, the situation is more variable across languages. As we already noted, the distal demonstrative determiner *nei* in Chinese is less restricted than its counterpart in the other languages since it only requires that its referent be identifiable, but not necessarily familiar. The languages also differ with respect to which demonstrative determiners, if any, require that the referent be activated. In Korean and Spanish, which have a three-way distinction in demonstratives, activation is a necessary condition only for the proximal determiner. Thus the medial as well as the distal demonstrative determiner is possible in (19) and (20).[6]

(19)   *Na-nun cinan pam   cam-ul cal-su-ka   æp-æss-ta.*
        I-TM   last   night   sleep-OM sleep-could-   neg-past-dec

        yæp cip $\begin{Bmatrix}k\bar{u}\\c\partial\end{Bmatrix}$ *kae-ka cam-ul   mot-ca-ke   hæe-ss-ta.*
        next house that   dog-SM sleep-OM neg-sleep-caus do-past-dec
        "I couldn't sleep last night. **That dog** next door made me not sleep"

(20)   *No   pudo dormir   a noche.*
        NEG could sleep   last.night

        $\begin{Bmatrix}Ese\\Aquel\end{Bmatrix}$ *perro de al   lado   no   me   dejó   dormir*
        that   dog   of to.the next.door NEG clitic let   sleep
        "I couldn't sleep last night. **That dog** next door kept me awake."

But in Japanese, which also has three demonstratives, both the proximal and medial determiners, like the demonstrative pronouns, require activation. Thus, although the facts are not entirely clear, the medial demonstrative *sono* appears to be inappropriate in an example like (21), where the dog can be assumed to be familiar but not activated.[7]

(21)   *Kinoo   wa   hitobanjuu remurenakatta*
        yesterday TM   all.night   couldn't.sleep

        $\begin{Bmatrix}?Sono\\Ano\end{Bmatrix}$ *inu no sei de*
        that   dog   reason
        "I couldn't sleep last night. **That dog** is the reason."

While the languages with three demonstratives differ as to whether the distal determiner alone (Japanese) or the medial as well as the distal determiner (Korean and Spanish) require that the referent be familiar. In all three languages as well as in Chinese, English and Arabic, the proximal demonstrative determiner requires that its referent be activated. Russian, however, differs from the other languages in lacking the activation condition on the proximal determiner. Thus either the proximal *eta* or the distal *ta* is possible in (22), even if the dog has not been mentioned in the immediate discourse context, i.e. if the dog is familiar, but not activated.

(22) $\begin{Bmatrix}Eta\\Ta\end{Bmatrix}$ *sobak u   soseda   menja*
        $\begin{Bmatrix}This\\That\end{Bmatrix}$ dog   at   neighbor me

        vsju noch' ne   davala spat'
        all   night not allow   to.sleep
        "**That dog** next door kept me awake all night."

Finally, the languages appear to differ in whether one or more demonstrative forms require not only that the referent be activated, but that it be speaker

activated. We noted that in English both pronominal and determiner *this* require speaker activation. Speaker activation also appears to be required for the proximal demonstratives *kono* and *kore* in Japanese and *i* in Korean. However, our data suggest that proximal demonstratives in the other languages don't have a speaker activation restriction.

Since the statuses on the givenness hierarchy are implicationally related, a status which meets necessary conditions for the use of a particular form will also meet necessary conditions for the use of all forms associated with statuses lower on the hierarchy. As we showed above for English, choice among forms when necessary conditions for the use of more than one form are met can, in many cases, be shown to follow from interaction of the givenness hierarchy with Grice's maxim of quantity. The final question we address, then, is whether the hierarchy interacts with the maxims in the same way across languages.

For all the languages we examined, we found, as we did for English, that if a referent was in focus the strongest possible form (i.e. a zero or unstressed pronominal) was generally used. This is shown in the tables in (23).[8]

(23)[9]

Arabic	focus	act	fam	unique
ø	44			
hua (sbj)	2			
hua (obj)	16			
daalika		3		
haadaa N	1	1		
daalika N		1	1	
al N	3	22	34	37

English	focus	act	fam	unique
it,he	38	1		
this N	2	1		
that N	2	2		
the N	12	26	23	22

Japanese	focus	act	fam	unique
ø	59			
kare		2		
kore	1	2		
sore		1		
sono		8		
kono N	1			
sono N	3	14		
ano N			1	
ø N	8	32	14	10

Chinese	focus	act	fam	unique
ø	35			
ta	23	4		
zhè		1		
zhè N	5	27		
nèi N		1	1	
ø N	9	12	14	17

Korean	focus	act	fam	unique
ø	20			
ku		4		
ku N	4	8	4	
ø N	4	2	11	2

Spanish	focus	act	fam	unique
ø	66	2		
él (sbj)	28	2		
él (obj)	32			
éste		1		
ése		6		
ese N		3	1	
el N	27	36	34	29

Use of a form that requires only activation generally implicates by Q1 that the referent is not in focus, i.e., it implies a focus shift , as in (24) and (25).

Chinese

(24) a. ... *xiao   haizi hen gaoxing suoyi ta ba ta dai   gei*
        small child very happy   so   he OM he bring give

    hong jinyu   de   yizhi hua   fang zai yugang limian.
    red   goldfish nom one   flower put in   bowl   loc

    "(The) small child was very happy. So he put a flower he had
    brought for (the) goldfish into (the) bowl."

   b. *zhe   jiu   shi zheige gushi*
      This   then be   this   story

    "**This** then is the (lit. this) story."

Japanese

(25) a. *toori e   dete   shibaraku   hashitteku*
        street to   go.out   for some time   run

    "He goes out onto the street and runs for some time."

   b. *to   nanka   yatai mitai na omise ga   atte de*
      and   something stall seem   shop SM   be   and

    "There is a shop like a stall."

   c. *kare wa   soko e sono omise no toko   e itte iku to*
      he   TM   there to that shop GM place to go   go   and

    "**He** goes to that shop."

Similarly, for the languages with an indefinite article, use of this form im-
plicates by Q1 that the referent is not uniquely identifiable and hence not familiar to
the addressee. Thus, the five languages we examined are similar to English in that
use of a particular form typically implicates that necessary conditions for a form
associated with a higher status don't obtain. That is, the form that gives most in-
formation about cognitive status is the one that is generally used

We also found, however, that for full definite NPs, Q2 rather than Q1 was
applicable in most contexts. As in English, when a full NP was used to refer to
something at least uniquely identifiable, it was generally the weakest possible form.
For Spanish and Arabic this was the definite article. For the other four languages,
which lack a definite article, a noun without a determiner was used in most contexts
where the referent was at least identifiable, and a demonstrative determiner in such
cases often serves as an explict signal to the addressee to search memory for a
familiar referent. Thus, the Russian example in (26) is analogous to the English
'reminder that's' discussed above.

(26)   *a soshli my na ostanovke kotoraja nazyvalas'*
        and went   we to stop   which   was.called

    sorok vtoroj kilometr a   spustilis' v   etot kan'on
    forty second kilometer and   went.down   in this canyon

    "And we went to a stop which was called forty second kilometer.
    And (we) went down into **that (lit. this) canyon**"

But if the referent of a full definite NP was activated or in focus, a demonstrative determiner was used more often in the languages which lack a separate definite article than in the languages which have a definite article. An example of such a use, from Japanese, is given in (27).

(27) *otokonoko ga modotte kite de sono otoko no*
      boy   SM return come and that male GM

      hito   tachi ga kake   o   yatte ano kingyo ga
      person pl.   SM gambling OM do   that goldfish SM

      torerutte   iu   koto ga wakatte
      be able to gain   say fact SM learn

      sorede sono otokonoko mo yaroo to suru n   desu ne
      then   that boy   also try   N   be   EM

      "After a while the boy comes back. He learns that those men are gambling to get that goldfish. And then that boy also gives it a try."

Demonstrative determiners thus occur more frequently in these languages than in English, Spanish and Arabic, since they are the only determiners that explicitly signal identifiability.

The research reported on here was not based on as large and varied a corpus as our previous work on English, so any conclusions we draw will necessarily be tentative. But it does appear that, despite obvious structural differences, the maxim of quantity interacts with the Givenness Hierarchy in a similar way across languages.

---

[1] A number of authors have proposed that different senses (or degrees) of givenness are correlated with different forms of reference (e.g. Chafe 1987, Gundel 1978a, 1978b, 1985, Lambrecht 1986, Prince 1981). See Gundel, Hedberg and Zacharski 1989 for discussion of differences between the Givenness Hierarchy and Prince's 1981 Familiarity Scale.

[2] To the extent that the syntactic and prosodic form of an utterance encodes its topic-commment structure and the relative importance of its constituents, membership in the in-focus set is partially determined by linguistic form. However, the actual inclusion of elements in the set depends ultimately on pragmatic factors, and is therefore not uniquely determinable from the syntax. See Gundel, Hedberg and Zacharski 1989 for further discussion.

[3] Since familiarity is the most common basis for identifiability, the application of Q2 here is predicted by the claim that the second part of the quantity maxim induces stereotypical interpretations (cf. Atlas and Levinson 1981, Horn 1984 and Levinson 1987). Thus, while some researchers consider familiarity to be part of the conventional meaning of the definite article and treat the non-familiar cases as exceptional (cf. Heim 1982) , we propose that only identifiability is conventionally signalled by the definite article and that familiarity is conversationally implicated by Q2.

[4] The methodology here was similar to that of the Pear Stories (see Chafe 1980). Speakers viewed a silent film, called the *The Golden Fish*, and, immediately after viewing the film, described it to another native speaker of their language.

[5] We are only talking about cognitive status here; there are, of course, other conditions such as structural restrictions on the distribution of personal pronouns vs. reflexives and restrictions on the use of definite determiners with generics and proper names which differ across languages.

[6] Although both forms require only familiarity, it is not the case that they are equally appropriate in all contexts. For the languages which have two demonstratives that require only familiarity, the more distal form is generally restricted to contexts where distance (either spatial or cognitive) is

being emphasized. Thus, Spanish speakers find the medial *ese* to be more natural than the distal *aquel* in an example like (20).

[7] There has been considerable debate among Japanese linguists regarding conditions on the appropriate use of *sono* and *ano* (cf. Hinds 1973, Kuno 1973, Kitagawa 1979 and Kuroda 1965). Our claim that *sono* requires activation whereas *ano* requires only familiarity appears to be at least consistent with all positions. But it is not clear whether this distinction is sufficient for explaining all differences in the distribution of these two forms.

[8] While all the languages but English are 'pro-drop', we find some interesting differences among them in the distribution of pronoun vs. zero. Chinese, Japanese, Korean and Russian allow zero in all argument positions (excluding objects of prepositions). But while the overwhelming majority of referents in focus were coded with zero in Japanese and Korean, almost half of the referents that were in focus in Chinese and Russian were coded with overt pronouns. Arabic and Spanish, on the other hand, allow zero in subject position only. But in Arabic, zero was used in almost all possible contexts (i.e. almost all in focus subjects were zero), whereas in Spanish (contrary to the often-stated claim that subject pronouns are used only for emphasis) we find pronouns used in almost half the contexts where zero would have been possible. The Japanese and Korean facts are predicted by our analysis since only a zero argument requires the referent to be in focus in these languages (i.e. Japanese and Korean have no unstressed pronouns, as seen in table 13). The facts in the other languages are consistent with our analysis, but we have no expanation for the differences in frequency of zero vs. overt pronoun.

[9] Table abbreviations are as follows: focus = in focus, act = activated, fam = familiar, unique = uniquely identifiable.

Atlas, Jay D. and Steven C. Levinson. 1981. It-clefts, informativeness and logical form. Radical Pragmatics, ed. by P. Cole, 1-61. New York: Academic Press.

Butt, John and Carmen Benjamin. 1988. A new reference grammar of Modern Spanish. London: Edward Arnold.

Chafe, Wallace L., ed. 1980. The pear stories: cognitive, cultural, and linguistic aspects of narrative production. Norwood, NJ: Ablex Publishing.

Chafe, Wallace. 1987. Cognitive constraints on information flow. Coherence and grounding in discourse, ed. by R. Tomlin, 21-51. Amsterdam: John Benjamins.

Fillmore, Charles J. 1975. Santa Cruz lectures on deixis. Bloomington: Indiana University Linguistics Club.

Fillmore, Charles. 1982. Toward a descriptive framework for spatial deixis. Speech, place, and action, ed. by R. J. Jarvella and W. Klein, 31-59. Chichester, NY: John Wiley & Sons.

Grice, H.P. 1975. Logic and conversation. Speech acts, ed. by P. Cole and J. Morgan, 41-58. New York: Academic Press.

Gundel, Jeanette K. 1978a. Stress, pronominalization and the given-new distinction. Department of Linguistics, University of Hawaii: University of Hawaii Working Papers in Linguistics 10(2).1-13.

Gundel, Jeanette K. 1978b. A universal constraint on deletion. Proceedings from the XIIth International Congress of Linguists, ed. by W. U. Dressler and W. Meid, 532-536. Innsbrucker Beiträge zur Sprachwissenschaft.

Gundel, Jeanette K. 1985. Shared knowledge and topicality. Journal of Pragmatics 9.83-107.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1988. On the generation and interpretation of demonstrative expressions. Proceedings of the 12th International Conference on Computational Linguistics, Budapest, 216-221.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1989. Givenness, implicature, and demonstrative expressions in English discourse. Papers from the Parasession on Language in Context, ed. by R. Graczk, B. Music, and C. Wiltshire. Chicago: Chicago Linguistic Society.

Hawkins, John. 1978. Definiteness and indefiniteness: a study in reference and grammaticality prediction. Atlantic Highlands, NJ: Humanities Press.

Heim, Irene R. 1982. The semantics of definite and indefinite noun phrases. Amherst: University of Massachusetts dissertation.

Hinds, John. 1973. Anaphoric demonstratives in Japanese. Journal of the Association of Teachers of Japanese 2.1-14.

Horn, Laurence R.. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature in context. Meaning, form, and use in context, ed. by D. Schiffrin, 11-42. Washington: Georgetown University Press.

Hwang, Shin Ja Joo. 1987. Discourse features of Korean narration. Dallas: Summer Institute of Linguistics.

Kitagawa, Chisato. 1979. A note on *sono* and *ano*. Explorations in linguistics: Papers in honor of Kazuko Inoue, ed. by G. Bedell, E. Kobayashi and M. Muraki, 232-243. Tokyo: Kenkyusha.

Kuno, Susumu. 1973. The structure of the Japanese language. Cambridge: The MIT Press.

Kuroda, S.-Y. 1965. Generative grammatical studies in the Japanese language.

Lakoff, Robin. 1974. Remarks on this and that. Papers from the Tenth Regional Meeting of the Chicago Linguistics Society, ed. by M. LeGaly, R. Fox and A. Bruck, 345-356. Chicago: Chicago Linguistics Society.

Lambrecht, Knud. 1986. Topic, focus, and the grammar of spoken French. Berkeley: University of California dissertation.

Levinson, Stephen C. 1987. Minimization and conversational inference. The pragmatic perspective, ed. by J. Verschueren and M. Bertuccelli-Papi, 61-129. Amsterdam: John Benjamins.

Li, Charles N. and Sandra A. Thompson. 1981. Mandarin Chinese: a functional reference grammar. Berkeley: University of California Press.

Martin, Samuel E. 1975. A reference grammar of Japanese. New Haven: Yale University Press.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. Radical Pragmatics, ed. by P. Cole. New York: Academic Press.

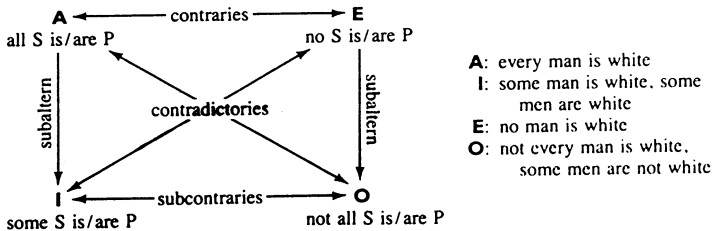# Hamburgers and Truth: Why Gricean explanation is Gricean[1]

Laurence R. Horn
Yale University

Since Dick Grandy (this volume) offers a review of the last twenty years of Gricean cooperation, I shall concentrate on the first 2300 years and focus—in the spirit of McGarrigle (to appear) - on the influence of Grice on Aristotle. It is by now widely accepted that the Gricean mechanism for the generation of generalized conversational implicatures through the exploitation of the maxim of quantity *(Make your contribution as informative as required for the current purposes of the talk–exchange)* provides a natural account of weak scalar operators *(some, possible, permitted, or, warm)* as semantically one-sided, lower–bounded by their literal meaning *(= at least some, at least possible,...)* with the two-sided understanding *(= some but not all, possible but not necessary,...)* derived by an upper– bounding scalar implicature:

(1)

	1-SIDED READING ->	2-SIDED READING
Max has 3 children.	'...at least 3...'	'...exactly 3...'
You ate SOME of the cookies.	'...some if not all...'	'...some but not all...'
It's POSSIBLE she'll win.	'...at least ◊...'	'...◊ but not certain...'
John is patriotic OR quixotic.	'...and perhaps both'	'...but not both'
It's WARM out.	'...at least warm...'	'...but not hot...'

This line provides a straightforward means for reconciling the apparent two-millenium-old conflict between the mutual implications intuitively relating the members of subcontrary pairs *(some/some not, possible/possible not,...)* and the desiderata of logical consistency and parsimony that remain unattainable when these implications are treated in semantic terms. But the quantity–driven pragmatic model of subcontrariety did not spring forth fully armed from the Gricean brow. I shall begin with a quick tour of some the highs and lows in the long history of subcontrariety, concentrating on those models that tend to prefigure the (neo–)Gricean approach to the Square of Opposition. The traditional assertoric square is laid out in (2) and the relevant terms of opposition identified in (3).

(2)

A: every man is white
I: some man is white, some men are white
E: no man is white
O: not every man is white, some men are not white

(3)a. Corresponding **A** and **E** statements are CONTRARIES and cannot be simultaneously true (though they may be simultaneously false).
 b. Corresponding **A** and **O** (and **I** and **E**) statements are CONTRADICTORIES; members of each pair cannot be simultaneously true OR simultaneously false.
 c. An **I** statement is the SUBALTERN of its corresponding **A** statement (and **O** of **E**); a subaltern is unilaterally entailed by its corresponding superaltern.
 d. Corresponding **I** and **O** statements are SUBCONTRARIES and cannot be simultaneously false (though they may be simultaneously true).

The last of these oppositions has a rich and turbulent history, beginning with Aristotle's recognition that while contraries and contradictories cannot hold simultaneously, 'the contradictories of a pair of contraries can sometimes be true with reference to the same subject; for instance, "not every man is white" and "some men are white" are both true' (*De Interpretatione* 17b23). Indeed, this is an 'opposition' in name only:

> Verbally four kinds of opposition are possible, viz. universal affirmative to universal negative [A/E], universal affirmative to particular negative [A/O], particular affirmative to universal negative [I/E], and particular affirmative to particular negative [I/O]: but really there are only three: for the particular affirmative is only verbally opposed to the particular negative. (*Prior Analytics* 63b21, emphasis added)

It was five centuries later before the square of opposition came along, and with it the topographic term SUBCONTRARY for this relation: the subcontraries appear beneath the contraries.

Modal values can be superimposed onto the same square of opposition, with the **A, I, E,** and **O** vertices assigned respectively to the necessary, the possible, the impossible, and the not necessary (possibly not). But for Aristotle the modal subcontraries are not only mutually compatible, they are - on what he calls the 'two-sided' reading of possible - equivalent. But if whatever is necessary is possible and whatever is possible is possibly not (not necessary), then whatever is necessary is not necessary (*De Int.*, Chapter 13). While Aristotle could have restricted the subalternation to one-sided possibility and subcontrariety to the two-sided variety, retaining logical consistency if not parsimony, he did not adhere to this distinction within his modal syllogistic, of which McCall (1963: 1) notes that 'perhaps no other piece of philosophical writing has had such consistently bad reviews'.

On the standard logical account of the subcontraries, particularity and possibility are treated as parallel and unambiguous, but only at the cost of ignoring the intuition that led Aristotle to the formulation of complementary conversion between *possible* and *possible not.* Just as *Some S is P* has been regarded (since Aristotle) as true so long as AT LEAST one S is P, so too *S may be P* or *It is possible for S to be P* has been taken (since Aristotle's disciple Theophrastus[2]) to be true provided it is AT LEAST possible for S to be P; *some* is compatible with *all* and *possible* with *necessary.* The 'one-sided' versions of both operators have thus won the day, while their 'two-sided' competitors *(some but not all, possible but not necessary)* have been relegated to the role of secondary, composite operators, when they are mentioned at all.

This approach has proved to be especially popular for the general assertoric statements, where a millenium of logicians have followed Avicenna's lead:

> If you say "some men are so-and-so", it is not necessary that some others are not so-and-so. If the proposition is about all, it is also about some. (Avicenna (ibn-Sina)/Zabeeh 1971: 24)

Nevertheless, there is an equally longstanding, if less hallowed, tradition of taking *some* to be two-sided and thus incompatible with *all.* Some have read it into Aristotle -

> On the Aristotelian theory...wherever the affirmative "some are" applies, the negative proposition "some are not" holds also. (Dewey 1938: 182)

> Aristotle seems to think that the main function of a particular statement is to describe a situation where the corresponding universal statement is false. His reasoning seems to be: If the universal is true, why assert the particular? (Rose 1968: 41)

But this reading appears dubious for Aristotle's assertorics, given his endorsement of the one-way subaltern entailment from **A** to **I** and from **E** to **O**:

> For having shown that it belongs to all, we will have shown also that it belongs to some; similarly, if we should show that it belongs to none, we will have shown also that it does not belong to all. (*Topics* 109a3)

Priority evidently belongs instead to the 5th-6th century Buddhist logician Diṅnāga and his colleagues who, in their *hetu-cakra* or Wheel of Reasons,

> do not admit four kinds of proposition like Aristotle and the Scholastics, but only three, since they interpret 'Some S is P' not as 'at least some' but as 'at least some and not all'...This would give a logical triangle in place of the western logical square. (Bochenski 1961: §53.14; cf. Tucci 1928)

This triangle of oppositions did not surface in the West until the mid-nineteenth century, when Sir William Hamilton of Edinburgh inaugurated a debate over the proper treatment of the subcontraries. Distinguishing two senses of *some*, the INDEFINITE *(at least some)* and the SEMI-DEFINITE *(some but not all)*, Hamilton (1860: 254) regarded the latter as basic: 'Some, if not otherwise qualified, means some only - this by presumption.' On this reading of the particular, the two statements *Some men are learned* and *Some men are not learned* are not only (as for Aristotle) compatible, given that their conjunction is consistent, but logically indistinct. The purported opposition between the two subcontraries, charged Hamilton (1860: 261), was 'only laid down from a love of symmetry, in order to make out the opposition of all the corners in the square of opposition...In reality and in thought, every quantity is necessarily either all, or none, or some. Of these the third...is formally exclusive of the other two.'

But even Hamilton tended to restore the indefinite *some* to its traditional place in his version of the syllogistic, although his practice was inconsistent enough to result in total incoherence, as his arch-rival Augustus De Morgan was quick to point out. While acknowledging the existence (at least in 'common language') of Hamilton's 'presumption' whereby *some* conveys *not all (some not)*, De Morgan defends the standard practice of relegating this inference to an extra-logical domain. De Morgan's subtle views are sampled below:

> IN COMMON CONVERSATION the affirmation of a part is meant to imply the denial of the remainder. Thus, by 'some of the apples are ripe', it is always [sic] intended to signify that some are not ripe. (De Morgan 1847: 4)

> *Some*, in logic, means *one or more, it may be all.* He who says that *some are*, is not to held to mean *the rest are not.* 'Some men breathe'...would be held false in COMMON LANGUAGE [which] usually adopts the complex particular proposition and implies that some are not in saying that some are. (De Morgan 1847: 56)

COMMON LANGUAGE makes a certain conventional approach to definiteness, which has been thrown away in works of logic. 'Some' usually means a rather small fraction of the whole; a larger fraction would be expressed by 'a good many'; and somewhat more than half by 'most', while a still larger proportion would be 'a great majority' or 'nearly all'.   (De Morgan 1847: 58)

With logicians the word *some* has in all time been no more than a synonym of *not–none:* it has stood for *one or more, possibly all.* WITH THE WORLD AT LARGE it is sometimes *possibly all,* sometimes *certainly not all,* ACCORDING TO THE MATTER SPOKEN OF. But with the logician "some are" is meerly and no more than the contradictory of "none are". Of these two one is true and the other false, and *some* equally contains some–certainly–not–all and some–possibly–all.
                                                            (De Morgan 1861: 51)

There are three ways in which one extent may be related to another...: complete inclusion, partial inclusion with partial exclusion, and complete exclusion. This trichotomy would have ruled the forms of logic, IF HUMAN KNOWLEDGE HAD BEEN MORE DEFINITE...As it is, we know well the grounds on which predication is not a trichotomy, but two separate dichotomies.   (De Morgan 1858: 121)

De Morgan's views are reflected in John Stuart Mill's even more directly proto-Gricean account of the subcontraries. In spurning Hamilton's innovations, Mill objects that

No shadow of justification is shown...for adopting into logic A MERE SOUS-ENTENDU OF COMMON CONVERSATION in its most unprecise form. If I say to any one, "I saw some of your children today", he might be justified in inferring that I did not see them all, NOT BECAUSE THE WORDS MEAN IT, but because, if I had seen them all, it is most likely that I should have said so: EVEN THOUGH THIS CANNOT BE PRESUMED UNLESS IT IS PRESUPPOSED THAT I MUST HAVE KNOWN WHETHER THE CHILDREN I SAW WERE ALL OR NOT.   (Mill 1867: 501)

Emphasis in the De Morgan and Mill citations is added to reflect Grice's influence here. Notice especially the epistemic rider on quantity-based inferences: the use of a weaker predicate suggests (implicates) that FOR ALL THE SPEAKER KNOWS the stronger predicate on the same scale could not have been substituted salva veritate.

Mill's allusion to a tacit principle that requires the speaker to use the stronger *all* in place of the weaker *some* when possible, and to draw the corresponding inference when the stronger term is not used, is echoed even by one of Hamilton's would-be supporters:

Whenever we think of the class as a whole, we should employ the term All; and therefore when we employ the term Some, it is implied that we are not thinking of the whole, but of a part as distinguished from the whole—that is, of a part only.   (Monck 1881: 156)
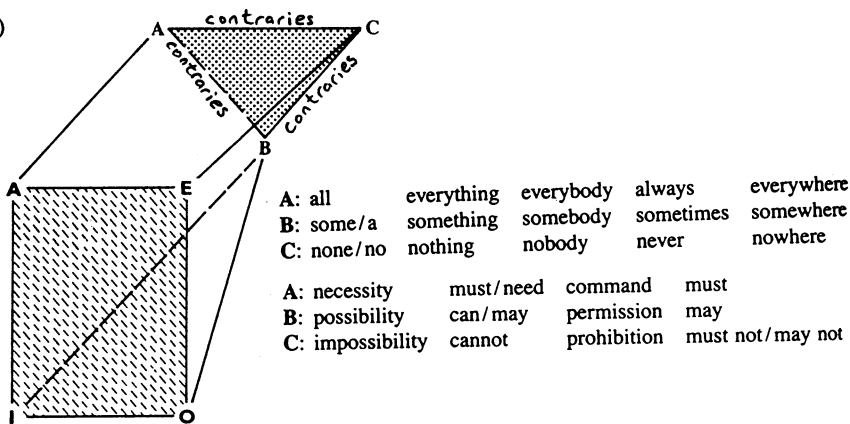
John Neville Keynes, the grandfather of modern economics, similarly noted in his 1906 *Logic* (202-3) that a speaker whose knowledge is incomplete cannot use *some S's are P* with the meaning 'some only'.

The idea that *some* should be assigned the two-sided rather than, or along with, the one-sided meaning did not die with Hamilton. Ginzberg (1913, 1914) carried the quarrel across the Channel, jettisoning the square of opposition for a triangle of contraries with vertices representing *all, none,* and *exactly some*—'quelques et rien que quelques'. But Couturat (1913, 1914), only too happy to play De Morgan to Ginzberg's Hamilton, attempts to dissuade his countryman from following 'le plus mauvais des logiciens' in collapsing the two distinct subcontraries into one basic proposition which is in fact a logical conjunction; he argues that the classical system cannot be perfected by adopting 'précisions' that are foreign to its very spirit.

The same logical triangle, still undrawn, makes an implicit reappearance in Jespersen's tripartition of logical operators (1917: Chapter 8). The category labels and instantiations here are Jespersen's, the geometry mine.

(4)

A:	all	everything	everybody	always	everywhere
B:	some/a	something	somebody	sometimes	somewhere
C:	none/no	nothing	nobody	never	nowhere

A:	necessity	must/need	command	must
B:	possibility	can/may	permission	may
C:	impossibility	cannot	prohibition	must not/may not

But while Jespersen's B category, the nadir of this Triangle of Opposition, corresponds SEMANTICALLY to the conjunction or neutralization of the I and O vertices of the traditional Square, it has the lexical membership of the I vertex *(some, possible)*. On logical, epistemic, and discourse grounds the identification of I and O is untenable, precisely for the traditional reason that the former is the contradictory of E, the latter of A.
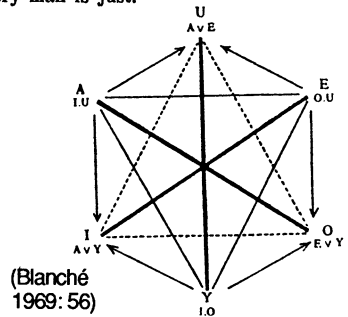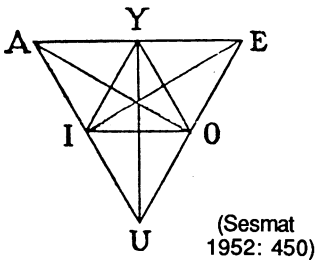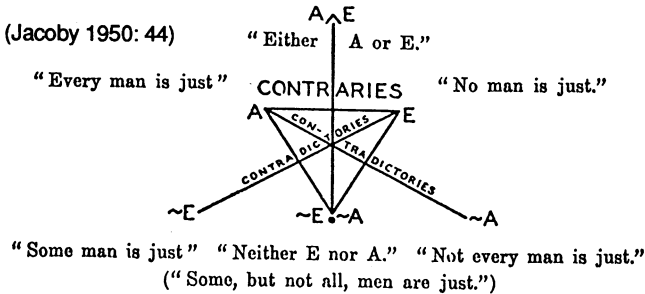
In fact, what we need here is not so much a triangle as a defective three-cornered square, given that in a wide variety of languages those values mapping onto the southeast corner of the square are systematically restricted in their potential for lexicalization. Thus alongside the quantificational determiners *all, some, no,* we never find an O determiner *nall;* corresponding to the quantificational adverbs *always, sometimes, never,* we have no *nalways* (= 'not always', 'sometimes not'). We may have equivalents for *both (of them), one (of them), neither (of them),* but never for *noth (of them)* (= 'not both', 'at least one...not'); we find connectives corresponding to *and, or,* and sometimes *nor* (= 'and not'), but never to *nand* (= 'or not', 'not...and')—at least not outside the lexicon of electronic circuitry. The missing O phenomenon, extending to the modals and deontics, is reinforced by a general tendency of O-->E drift, wherein lexical items or collocations associated by their compositional form or etymology with the O corner of the square move inevitably northward toward E. As I have argued more fully elsewhere (Horn 1972; Horn 1989: §4.5), the pragmatic inferential relation between the positive and negative subcontraries results in the superfluity of one of these subcontraries for lexical realization, while the functional markedness of negation assures that the superfluous, unlexicalized subcontrary will always be O.[3]

Thus Jespersen's tripartition gets the lexical facts right, but for the wrong reason. With characteristic insight, Sapir opts for a solution midway between the classical Square and the Jespersenian Triangle. His particular subcontraries are neither semantically bilateral nor strictly unilateral:

> 'Not everybody came' does not mean 'some came', which is implied, but 'some did not come'. Logically, the negated totalizer [not every] should include the totalized negative, i.e. opposite or contrary [none], as a possibility, but ordinarily this interpretation is excluded. (Sapir 1930: 21)
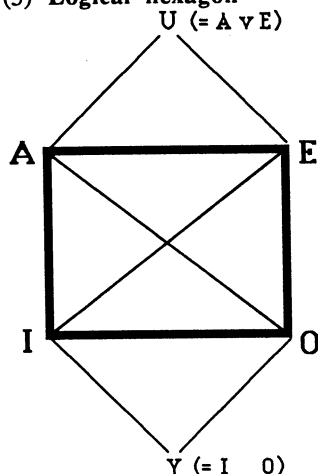
Note especially Sapir's use of *is implied* (vs. *means*) and his qualifier *ordinarily*, emphasizing the essential role of the context in licensing the implication in question.

One final triangulist salvo was launched in the early 1950's, appropriately enough by three philosophers working independently but exploiting essentially the same geometry. For Jacoby (1950), Sesmat (1951-2), and Blanché (1952, 1953, 1969), the square and triangle of (4) can be combined to form a logical hexagon on which the diametrically opposed terms are contradictories:
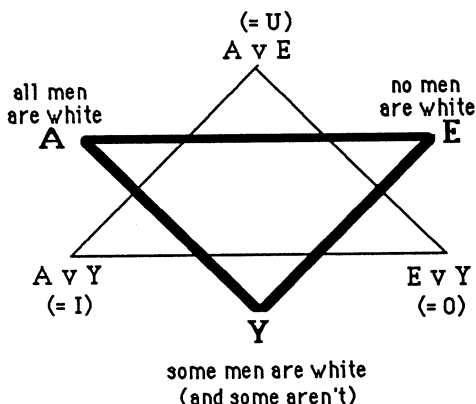

(Jacoby 1950: 44)

"Either A or E."

"Every man is just"   CONTRARIES   "No man is just."

"Some man is just" "Neither E nor A." "Not every man is just."
("Some, but not all, men are just.")

(Sesmat 1952: 450)

(Blanché 1969: 56)

More accurately, a major triangle of mutual contraries **A E Y**, representing De Morgan's trichotomy of definite human knowledge, is superimposed upon a minor triangle of subcontraries whose vertices - designated appropriately enough as **I O U**[4]—are disjunctively defined. Curiously, Doyle, Sesmat, and Blanché fail to note that what we want here is not so much the Logical Hexagon of (5) as the Logical Magen David of (6):

(5) Logical hexagon

(6) Logical Magen David

These efforts to redesign the square have met with general nonacclaim[5] or, in the case of Jacoby's (1950: 43-44) 'double triangle', with a prescient response by an unfortunately obscure proto-Grice in an equally obscure Jesuit journal:

> What can be understood without being said is usually, in the interest of economy, not said...A person making a statement in the form, "Some S is P", generally wishes to suggest that some S also is not P. For, in the majority of cases, if he knew that all S is P, he would say so...If a person says, "Some grocers are honest", or "Some books are interesting", meaning to suggest that some grocers are not honest or that some textbooks are not interesting, he is really giving voice to a conjunctive proposition in an elliptical way.
>
> Though this is the usual manner of speech, there are circumstances, nevertheless, in which the particular proposition should be understood to mean just what it says and not something else over and above what it says. One such circumstance is that in which the speaker does not know whether the subcontrary proposition is also true; another is that in which the truth of the subcontrary is not of any moment.
>
> (Doyle 1951: 382)

Thus, a host who has received a couple of acceptances and no declinations from his invitees could felicitously claim that some of those invited will come without licensing an inference from I to O and hence to Y. In more current terminology, Doyle depicts Quantity as potentially overridden by either Quality or Relation (cf. Horn 1984 and below). But like De Morgan, Mill, and Monck before him, Doyle must tacitly appeal to a crucial principle yet to be explicitly codified.

An early formulation of this principle is offered by Strawson (1952: 178-9), who however credits his 'general rule of linguistic conduct' to 'Mr. H. P. Grice':

> One should not make the (logically) lesser, when one could truthfully (and with greater or equal clarity) make the greater claim.

Grice's own 'first shot' at the relevant principle (1961: 132)—

> One should not make a weaker statement rather than a stronger one unless there is a good reason for so doing.

—later evolves into his [FIRST] MAXIM OF QUANTITY (Grice 1967/1975: 45):

> Make your contribution as informative as is required (for the current purposes of the talk–exchange).

Fogelin arrives independently at the same principle in the form of his RULE OF STRENGTH (1967: 20-22):

> Make the strongest possible claim that you can legitimately defend!

Invoking a Gricelike 'distinction between what a statement implies (or entails) and what the use of a statement indicates', Fogelin emphasizes the connection between the rule-governed nature of language and the license to draw inferences obtained through the assumption that rules are being obeyed:
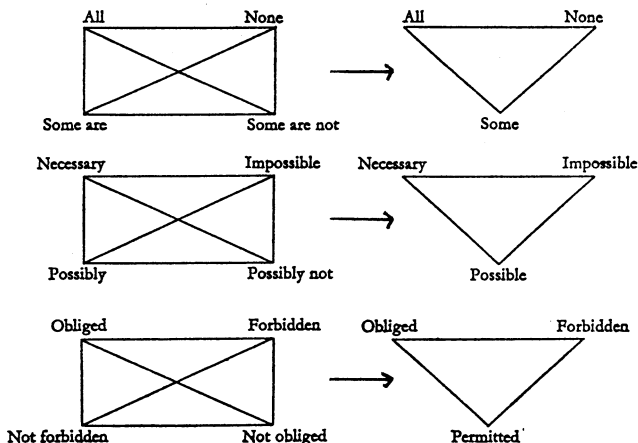
> The use of language is under the governance of rules and thus when someone employs a given expression we are entitled to assume that the appropriate rules are being followed. When we can draw inference from the use of a statement that we cannot draw from the statement itself, this usually indicates that our inference is grounded on the assumption that some linguistic rule is in force.

In the case of the Rule of Strength, we have these corollaries:

> •Do not employ an **I** or an **O** proposition in a context where you can legitimately employ an **A** or an **E** proposition...The *use* of one subcontrary typically suggests the appropriateness of using the other.
> •Do not affirm one subcontrary if you are willing to deny the other.
> • Subcontraries tend to collaspe together.

Thus accoutered, Fogelin (1967: 22) tries his own hand at beating squares into triangles:

(7)

All   None   All   None

Some are   Some are not   Some

Necessary   Impossible   Necessary   Impossible

Possibly   Possibly not   Possible

Obliged   Forbidden   Obliged   Forbidden

Not forbidden   Not obliged   Permitted

As our earlier avatars from De Morgan to Doyle recognized, the arrows are activated only when the context allows. Thus Fogelin's triangles - as distinct from those of the Jacoby-Sesmat-Blanché triumvirate - are pragmatically derived and not semantically driven.

Finally, we come to Harnish's MAXIM OF QUANTITY-QUALITY (1976: 362), Make the strongest relevant claim justifiable by your evidence, which decomposes into three subrules at potential loggerheads: Be as informative as necessary, Be relevant, Have evidence for what you say. Harnish cites Grice and Fogelin, as well as O'Hair (1969:45):

> Unless there are outweighing good reasons to the contrary, one should
> not make a weaker statement rather than a stronger one if the audience is
> interested in the extra information that would be conveyed by the latter.

O'Hair in turn takes Grice 1961 as a starting point, while Fogelin, who takes no cognizance of Grice, leans on Nowell-Smith, who seems to have been equally unaware of the existence of his erstwhile fellow Oxonian Paul Grice, and vice versa.[6] Nowell-Smith's definition and rules of contextual implication (1954: 80-82) are given as follows:

> A statement *p* contextually implies a statement *q* if anyone who knew the
> normal conventions of the language would be entitled to infer *q* from *p*
> *in the context in which they occur.*
> Rule 1: When a speaker uses a sentence to make a statement, it is contextually
> implied that he believes it to be true.   [cf. Grice's Quality$_1$]
> Rule 2: A speaker contextually implies that he has what he himself believes to
> be good reasons for his statement.   [cf. Grice's Quality$_2$]
> Rule 3: What a speaker says may be assumed to be relevant to the interests of
> his audience.   [cf. Grice's Relation]

Clearly an idea whose time had come. We have in embryo here all of Grice's content maxims EXCEPT Quantity or Strength; by putting Nowell-Smith together with Fogelin we arrive in the neighborhood of Grice (1967/1975: 45-6) and William James:

QUALITY: Try to make your contribution one that is true.
1. Do not say what you believe to be false.
2. Do not say that for which you lack evidence.
QUANTITY:
1. Make your contribution as informative as is required
(for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.
RELATION: Be relevant.
MANNER: Be perspicuous.
1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief. (Avoid unnecessary [sic] prolixity.)
4. Be orderly.

But while Nowell-Smith discusses irony, lying, and play-acting as 'secondary uses of language', he offers no general account of speaker meaning, cooperation, or exploitation to explain how conveyed meaning arises, just as Mill, Doyle, and Fogelin, despite their recognition of a quantity or strength rule exploitable to generate the pragmatic upper-bounding of the subcontraries, lack a coherent set of rules or maxims whose interaction yields the rich array of nonlogical inferences in context described in the post-Gricean literature. More specifically, the forerunners never explicitly anticipate Grice's Cooperative Principle—Make your conversational contribution such as is required, at the state at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged—of which (as Georgia Green points out in her contribution to this volume) the maxims must be seen merely as special instances. It was Paul Grice who put it all together.

The invocation of the maxims may not be universally regarded as a breakthrough, especially given their self-evident or trivial appearance. The skeptic may recall Lord Macaulay's general maxim that nothing is so useless as a general maxim, and it might be further wondered whether Grice's neo-Kantian gang of four (or nine, counting submaxims) represents a significant advance over the rather bulkier set assumed by an earlier colonist also commemorated this BLS weekend (Washington [1746]1988). My proposed mapping relations between Grice's conversational maxims and George Washington's 'rules of civility and decent behavior in company and conversation' would not necessarily be endorsed by the General, wary as he notoriously was of foreign entanglements.

(8)   *79th:* Be not apt to relate news if you know not the truth thereof.
[cf. Quality]
      35th: Let your discourse with men of business be short and comprehensive.
[cf. Quantity]
      90th: Being set at meat, scratch not; neither spit, cough, or blow your nose,
except if there is a necessity for it.   [cf. Relation]
      73th: Think before you speak; pronounce not imperfectly nor bring out your
words too hastily, but orderly & distinctly.   [cf. Manner]
      80th: Be not tedious in discourse or in reading unless you find the company
pleased therewith.   [cf. Manner]

The appeal to informativeness or strength in the various castings and recastings of the principle invoked implicitly by Mill, Monck, and Doyle, and explicitly since Grice, assumes that such a notion can be defined and quantified. An obvious starting point here is the relation of unilateral entailment or proper inclusion of classes, as recognized by Fogelin: 'A proposition *"a"* is stronger than a proposition *"b"* if *"a"* implies *"b"* but *"b"* does not imply *"a"* '—for some appropriately defined sense of 'implies' stronger than
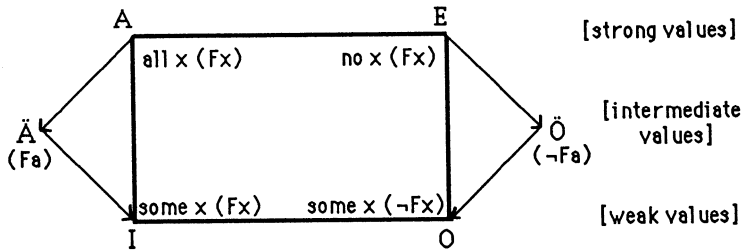
material implication (Fogelin 1967: 20). A similar conception of informativeness can be discerned in Aristotle:

> If one is to say of the primary substance what it is, it will be more informative and apt to give the species than the genus. For example, it would be more informative to say of the individual man that he is a man than that he is an animal (since the one is more distinctive of the individual man while the other is more general); and more informative to say of the individual tree that it is a tree than that it is a plant.
>
> (Categories 2b10ff.)

Of course, as recognized by Fogelin, and no doubt by Aristotle and Grice, to provide a real (and not just working) definition of informational strength is no simple matter. Some of the difficulties involved are addressed in Thomason 1987 and Ginzburg 1989.

But strength is not enough. In (5) above we extended the Square of Opposition vertically to produce a tall hexagon by including a southerly **Y** vertex corresponding to the conjunction of the subcontraries and a northerly **U** vertex corresponding to the disjunction of the contraries. But we can also follow the (mutually independent) suggestions of Czezowski (1955) and Fogelin (1967: 17) and extend the Square OUTWARD to form a FAT hexagon. This move is motivated by the need to represent singular propositions with respect to their universal and particular counterparts, unmodalized propositions with respect to the necessary and the possible, and so on. Adopting Ä for the intermediate positive (westerly) value between **A** and **I**, and Ö for its contradictory, the negative (easterly) intermediate value between **E** and **O**, we get the figure below:

The umlauted values are intermediaries of subalternation in that each unilaterally entails the weak value below it and is unilaterally entailed by the strong value above it. Thus *Everybody won* unilaterally entails *Kim won,* which in turn unilaterally entails *Somebody won;* if nobody came then Lee didn't, and if Lee didn't then someone didn't. Similarly, if Kim won AND Chris won (A) then Kim won (Ä), and if Kim won then Kim won OR Chris won (I); *necessarily* p (A) unilaterally entails p (Ä) which entails *possibly* p (I), and so on for all sets of values mapping onto the positive and negative sides of the fat hexagon.

But the symmetry of the fat hexagon belies an asymmetry in implicature. If I tell you that my wife is either in the kitchen or the bedroom, you will infer that I don't know that she's in the kitchen (Grice 1961: 130). But I can inform you that the kitchen is a mess without implicating that the bedroom isn't. If you tell me something is POSSIBLY true, I will assume you don't know it's true, but if you tell me that something is true (e.g. that all bachelors are unmarried), I will not assume you don't know it's NECESSARILY true. That is, the use of the weak I or O form proposition licenses the inference that the speaker was not in a position to use the corresponding intermediate (or strong) proposition, but the use of an intermediate Ä or Ö form does NOT quantity-implicate the negation of its strong

> If I were to say to you, 'My sister is coming to lunch tomorrow', I do presuppose that I have a sister but in presupposing it I do not necessarily assume that you have a prior assumption or belief that I have a sister.
>
> (Burton-Roberts 1989: 26)

From this observation, Burton-Roberts correctly concludes that presupposition cannot be defined directly in terms of mutual knowledge. But then nobody ever said it could. For Stalnaker 1974, pragmatic presuppositions are 'propositions whose truth [the speaker] takes for granted, or seems to take for granted, in making his statement' (1974: 198); presupposed material can be communicated as new information by a speaker who 'tells his auditor something...by pretending that his auditor already knows it' (Stalnaker 1974: 202). The linking of presuppositions to what is potentially non-controversial, rather than to what is mutually known, was motivated by an observation of Jerry Sadock:

> I am asked by someone who I have just met, "Are you going to lunch?" I reply, "No, I've got to pick up my sister." Here I seem to presuppose that I have a sister even though I do not assume that the speaker knows this.
>
> (Stalnaker 1974: 202, citing Sadock (p.c.))

The idea that a speaker can ACT AS IF a proposition is part of the common ground when it isn't, and thereby force the hearer to adjust her map of the common ground to encompass that presupposed proposition, was later codified in Lewis's RULE OF ACCOMMODATION for presupposition:

> If at time t something is said that requires presupposition P to be acceptable, and if P is not presupposed just before t, then—*ceteris paribus* and within certain limits—presupposition P comes into existence at t.
>
> (Lewis 1979: 340)

But this notion of accommodation, which Lewis generalizes to permission statements, descriptions, vagueness, relative modalities, performatives, and planning, and which Sperber & Wilson (1986) put to their own uses in Relevance Theory, is itself—as Stalnaker recognized—just a special case of Gricean exploitation. So we should not be surprised to find Grice himself (1981: 190) forging the same connection:

> It is quite natural to say to somebody...*My aunt's cousin went to that concert*, when one knows perfectly well that the person one is talking to is very likely not even to know that one had an aunt, let alone know that one's aunt had a cousin. So the supposition must be not that it is common knowledge but rather that it is noncontroversial, in the sense that it is something that you would expect the hearer to take from you (if he does not already know).

Grice, working in isolation from the Stalnaker-Lewis tradition—indeed, this 1981 publication recapitulates a 1970 talk in Urbana which precedes the development of that tradition—thus provides the foundation for a tenable construct of pragmatic presupposition. But unintentionally so: faithful to his Oxford roots, Grice himself always saw presupposition as a semantic notion. Of course, he didn't really NEED pragmatic presuppositions since his conventional implicatures (Grice 1975) fit the same job description, as Karttunen & Peters (1979) have shown. But even the Oxonian in charge of the presupposition detail was no stranger to accommodation, although it is not clear how any notion of presupposition so qualified can remain semantic. Here is Strawson (1950: 344) polishing off his shot across the Russellian bow:

counterpart, **A** or **E** respectively. Since there is no quantity- or information-based distinction between these (sub)subalternations, we must seek the source of the asymmetry elsewhere. As O'Hair (1969: 45-48) observes for the disjunctive cases, the crucial distinction here relates not to the content (what is said) but to the form (HOW what is said is said): it is because the intermediate values are not only more informative but BRIEFER than their more southerly counterparts that the use of the latter will strongly implicate against the former. But the strong values, while more informative than their umlauted mates, are more prolix, so Quantity here is offset by Manner and potentially by Relation. The richness of Grice's framework makes it possible to begin to develop a theory of not just what CAN be implicated but what WILL be implicated in a given context.

I have argued (Horn 1984, 1989) that if we assume (with Grice 1975: 46-7 and contra Sperber & Wilson 1986) that Quality is primary and essentially unreducible, we can attempt to boil the remaining maxims and submaxims down to two fundamental countervailing principles. Within the dualistic functional model I propose, the **R** Principle—MAKE YOUR CONTRIBUTION NECESSARY—is an upper-bounding speaker-oriented correlate of the Law of Least Effort dictating minimization of form: SAY NO MORE THAN YOU MUST (GIVEN **Q**). The **Q** Principle—MAKE YOUR CONTRIBUTION SUFFICIENT—is a lower-bounding hearer-based guarantee of the sufficiency of informative content: Say as much as you can (given **R** and the Maxim of Quality; cf. Doyle 1951: 382, cited above). **R** collects Gricean Relation, the second submaxim of Quantity[7], and the last two submaxims of Manner, while **Q** responds to the first submaxim of Quantity and the first two Manner submaxims. The functional tension between these two antinomic principles governs not just the determination of implicatures but a wide range of linguistic phenomena, from lexical change to politeness strategies, from periphrastic causatives to logical double negation, from euphemism to the interpretation of pronouns and gaps; cf. Horn 1984, 1989 for details and Sperber & Wilson 1986 and Levinson 1987 for other attempts to reduce or reconstruct the maxims and to predict the resolution of maxim clashes.

Grice's model of conversational interaction and nonlogical inference is most dramatically distinguished from competing accounts by his emphasis on how the exploitation of shared tacit principles allows an interlocutor to map what was SAID into what was MEANT based on what was NOT said. This feature is also present in the independently arrived-at pragmatic theory of Oswald Ducrot, which is equipped with its own version of the quantity or strength rule and its own definition of exploitation:

> [The *Loi d'exhaustivité*] exige que le locuteur donne, sur le thème dont il parle, les renseignments les plus forts qu'il possède, et qui sont susceptible d'intéresser le destinataire...Le destinataire, supposant que le locuteur a respecté cette règle, aura tendance, si la réserve du locuteur ne peut pas être attribuée à une absence d'information, à interpréter toute affirmation restreinte comme l'affirmation d'une restriction (s'il ne dit que cela, alors qu'il sait ce qui s'est passé, c'est qu'il n'y a que cela).
>
> (Ducrot 1972: 134)

But Grice's notion of exploitation, unlike Ducrot's, plugs into a comprehensive system of maxims and extends from conversational to conventional inferences in ways Grice himself may not fully recognize. In his recent defense of semantic presupposition, Noel Burton-Roberts points out that a pragmatic theory of presupposition framed 'in terms of assumption-sharing between speaker and hearer' is 'quite simply wrong':

A literal-minded and childless man asked whether all his children are asleep will certainly not answer "Yes" on the ground that he has none; but nor will he answer "No" on this ground. Since he has no children, the question does not arise. To say this is not to say that I may not use the sentence, "All my children are asleep" with the intention of letting some one know that I have children, or of deceiving him into thinking that I have. Nor is it any weakening of my thesis to concede that singular phrases of the form "the so-and-so" may sometimes be used with a similar purpose. Neither Aristotelian nor Russellian rules give the exact logic of any expressions of ordinary language; for ordinary language has no exact logic.

      The collapsing of Strawson's sleeping children into the sister of Stalnaker and Sadock, who herself metamorphoses into Grice's aunt's concert-going cousin, who in turn mutates into the lunch-going sister of Burton-Roberts, should remind us that in the evolution of pragmatic theory, all progress is relative.
      In the valuable Retrospective Epilogue to his collected works, Grice relates the genesis of the William James lectures. Having developed (though not yet named) the doctrines of exploitation and conversational implicature in response to Wittgensteinian objections to the causal theory of perception, he recalls (1989: 374-75),

> It then occurred to me that apparatus which had rendered good service in one area might be equally successful when transferred to another; and so I canvassed the idea that the alleged divergences between Modernists' Logic and vulgar logical connectives might be represented as being a matter not of logical but of pragmatic import.

For philosophers, the most significant of these divergences is that between the material conditional and the if-then of ordinary language. But Grice's Modified Occam's Razor ('senses are not to be multiplied beyond necessity'), honed with his personal philosopher's stone (the cooperative principle and its component maxims), cuts a wide swath through the ancient thickets of meaning and ambiguity—i.e. GRICE SAVES.
      Grice observes in the same retrospection (1989: 375) that 'when a sentence which used in isolation standardly carries a certain implicature is embedded in a certain linguistic context, for example appears within the scope of a negation-sign', that negative operator may 'be interpreted as governing not the conventional import but the nonconventional implicatum of the embedded sentence'. I have argued (Horn 1985) that both Grice's case in point, the denial of a conditional, and that of 'paradoxical negation' in scalar contexts *(You didn't eat v SOME of the cookies, you ate ALL of them)* can be subsumed within a generalized neo-Ducrotian theory of metalinguistic negation.
      But where, you may ask, do the hamburgers come in? Grice departs crucially from his predecessors and from such coevals as Fogelin and Ducrot in regarding linguistic cooperation in the conversational enterprise as a subcase of a general theory of rationality (see again Georgia Green's paper in this volume). Thus, he reminds us,

> It is irrational to bite off more than you can chew whether the object of your pursuit is hamburgers or the Truth.   (Grice 1989: 369)

Ever true to the spirit of Quantity and to Washington's 97th Maxim,

Put not another bit into your mouth till the former be swallowed. Let not your morsels be too big. (Washington 1988: 27)

Grice was always rational enough to bite off neither more nor less than his appetite allowed. But no man lives by meat alone, much less a philosopher of language large enough to bestride the warring camps of Russell's Modernists and Strawson's Neo-Traditionalists (Grice 1989: 372). And anyway hamburgers need rolls. So it is meet that such a healthy portion of the Gricean legacy on pragmatic inference consists not of solutions but of problems and questions, of roadmaps and menus. For, as Grice reminds us elsewhere in offering a defense of absolute value admittedly 'bristling with unsolved or incompletely solved problems',

If philosophy generated no new problems it would be dead... Those who still look to philosophy for their bread-and-butter should pray that the supply of new problems never dries up. (Grice 1986: 106)

Fortunate indeed are we linguistic philosophers and philosophical linguists, nourished with the ground meat of conversational logic and the fresh bread of Gricean analysis. We know that we shall never starve, for we have been served the biggest Mac of all.

Footnotes

[1] I would like to acknowledge the help of Jay Atlas, Bob Fogelin, Dick Grandy, Georgia Green, Jerry Sadock, Bob Stalnaker, and Jean Thomson in lighting my way along some of the longer and windier byways of scholarship.

[2] As Aristotle's pupil and successor as head of the Peripatetic school, as well as his nephew, the executor of his will, and the lover of his son (Suidas/Bekker 1854: 498), Theophrastus presumably knew whereof he spoke.

[3] That the O vertex, unlike its three square companions, enjoys no simple representation was recognized by St. Thomas, who observed that whereas in the case of the universal negative (A) 'the word "no" [nullus] has been devised [sic!] to signify that the predicate is removed from the universal subject according to the whole of what is contained under it', when it comes to the PARTICULAR negative (O), we find that

there is no designated word, but 'not all' [non omnis] can be used. Just as 'no' removes universally, for it signifies the same thing as if we were to say 'not any' [i.e. 'not some'], so also 'not all' removes particularly inasmuch as it excludes universal affirmation.
(Aquinas, in Arist. de Int., Lesson X, Oesterle 1962: 82-3)

[4] It will be noticed that Sesmat's hexagon has the Y above and the U below, as does the somewhat sketchier model of Hegenberg 1957. I opt here for Blanché's vowel system for its mnemonic value. Von Wright (1951) proposes in effect a logical pentagon, with a nadir (= our Y) for the conjunction of I and O but no apex (= U) for its contradictory.

[5] In fact, the triangulist perspective—minus the geometry—has its adherents still. Thus Kuroda (1977: 97-8) posits an 'every-day reading' of Some animals are white, which is 'assumed to entail' Some animals are not white, so that the two come out 'logically equivalent'. Kuroda is not dissuaded from this 'logical equivalence' by his recognition that

on its 'every-day reading', *Some animals are white* cannot serve as the contradictory of *No animals are white* since both propositions will be false if ALL animals are white. A semantic account of assertoric and modal subcontrariety is also endorsed by Morpurgo-Tagliabue (1981: 502):

> The ∃x, the 'possible', may, to some extent, come nearer and nearer to the 'all (x)', the 'necessary', without ever reaching it, like Achilles and the tortoise...It is excluded that while saying 'not-all' (O) one could mean 'nobody' and saying 'not-nobody' (I) one could mean 'all'...If I say "not all people are clever" (O), this means that there are some who are stupid.

[6]While something was clearly in the air in the Oxford of the early 1950's, the Oxonian atmosphere was decidedly diffuse. His exclusion from what Grice (1986: 49) fondly recalls as the ordinary language philosophers' 'Play Group' that met every Saturday until the death of their primus inter pares J. L. Austin may have placed Nowell-Smith out of the loop as far as the development of the theory of contextual implication was concerned.

[7]Note the connection between Relation and the second Quantity submaxim: what could make a speaker's contribution more informative than is required except the inclusion of material irrelevant for the current purposes of the exchange?

References

Aristotle. Categories and De interpretatione, ed. & trans. by J. L. Ackrill. Oxford: Clarendon Press, 1963. [Other citations from Oxford U. P. edition of complete works.]

Avicenna. Treatise on logic, ed. and trans. by F. Zabeeh. The Hague: M. Nijhoff, 1971.

Blanché, Robert. 1952. Quantity, modality, and other kindred systems of categories. Mind 61. 369-75.

Blanché, Robert. 1953. Sur l'opposition des concepts. Theoria 19. 89-130.

Blanché, Robert. 1969. Structures intellectuelles, deuxième éd. Paris: J. Vrin.

Bochenski, Innocentius Marie. 1961. A history of formal logic, trans. by Ivo Thomas. Notre Dame: University of Notre Dame Press.

Burton–Roberts, Noel. 1989. The limits to debate: A revised theory of semantic presupposition. Cambridge: Cambridge University Press.

Couturat, L. 1913. Des propositions particulières et de leur portée existentielle. Revue de Métaphysique et de Morale, March 1913, 256-59.

Couturat, L. 1914. Réponse [to Ginzberg 1914]. Revue de Métaphysique et de Morale, March 1914, 259-60.

Czezowski, Tadeusz. 1955. On certain peculiarities of singular propositions. Mind 64. 392-95.

De Morgan, Augustus. 1847. Formal logic. London: Taylor & Walton.

De Morgan, Augustus. 1858. On the syllogism: III, and on logic in general. On the syllogism and other logical writings, 74-146. London: Routledge & Kegan Paul.

De Morgan, Augustus. 1861. Hamiltonian logic. Athenæum 13 July 1861, p. 51.

Dewey, John. 1938. Logic: The theory of inquiry. New York: Henry Holt and Company.

Doyle, John J. 1951. In defense of the square of opposition. The New Scholasticism 25. 367-96.

Ducrot, Oswald. 1972. Dire et ne pas dire. Paris: Hermann.

Fogelin, Robert J. 1967. Evidence and meaning. New York: Humanities Press.

Gazdar, Gerald. 1979. Pragmatics. New York: Academic Press.
Ginzberg, S. 1913. Note sur le sens équivoque des propositions particulières. Revue de Métaphysique et de Morale, January 1913, 101-6.
Ginzberg, S. 1914. À propos des propositions particulières. Revue de Métaphysique et de Morale, March 1914, 257-9.
Ginzburg, Jonathan. 1989. Towards a more informed and informational view of informativeness. Ms., Stanford University.
Grice, H. P. 1961. The causal theory of perception. Proceedings of the Aristotelian Society, Supplementary Volume 35. 121-52.
Grice, H. P. 1967. Logic and conversation. William James lectures at Harvard.
Grice, H. P. 1975. Logic and conversation. Syntax and Semantics 3: Speech acts, ed. by Peter Cole and Jerry L. Morgan, 41-58. New York: Academic Press.
Grice, H. P. 1981. Presupposition and conversational implicature. Radical pragmatics, ed. by Peter Cole, 183-98. New York: Academic Press.
Grice, H. P. 1986. Reply to Richards. Philosophical grounds of rationality, ed. by Richard Grandy and Richard Warner, 45-106. Oxford: Clarendon Press.
Grice, H. P. 1989. Studies in the way of words. Cambridge: Harvard University Press.
Grote, George. 1872. Aristotle, Volume I. London: John Murray.
Hamilton, Sir William. 1860. Lectures on Logic, Volume I. Edinburgh: Blackwood.
Harnish, Robert M. 1976. Logical form and implicature. An integrated theory of linguistic ability, ed. by Thomas G. Bever et al., 464-79. New York: Crowell.
Hegenberg, Leonidas. 1957. A negação. Revista Brasieira de Filosofia 7. 448-57. Summarized and reviewed by Hector-Neri Castañeda, J. Symb. Logic 25. 265 (1960).
Hirschberg, Julia. 1985. A theory of scalar implicature. University of Pennsylvania dissertation.
Horn, Laurence. 1972. On the semantic properties of logical operators in English. UCLA dissertation. Distributed by Indiana University Linguistics Club, 1976.
Horn, Laurence. 1973. Greek Grice: a brief survey of proto–conversational rules in the history of logic. CLS 9, 205-214.
Horn, Laurence. 1984. Toward a new taxonomy for pragmatic inference: Q–based and R–based implicature. Meaning, form, and use in context: linguistic applications (GURT '84), 11-42. Washington: Georgetown University Press.
Horn, Laurence. 1985. Metalinguistic negation and pragmatic ambiguity. Language 61. 121-74.
Horn, Laurence. 1989. A natural history of negation. Chicago: U. of Chicago Press.
Jacoby, Paul. 1950. A triangle of opposition in Aristotelian logic. The New Scholasticism 24. 32-56.
Jespersen, Otto. 1917. Negation in English and other languages. Copenhagen: A. F. Høst.
Karttunen, Lauri. 1974. Presupposition and linguistic context. Theoretical Linguistics 1. 181-93.
Karttunen, Lauri and Stanley Peters. 1979. Conventional implicature. Syntax and Semantics 11: Presupposition, ed. by Choon-Kyu Oh & David A. Dinneen, 1-56. New York: Academic Press.
Keynes, John Neville. 1906. Studies and exercises in formal logic. London: Macmillan.
Kuroda, Sige-Yuki. 1977. Description of presuppositional phenomena from a non–presuppositionist point of view. Lingvisticæ Investigationes 1. 63-162.
Levinson, Stephen C. 1987. Minimization and conversational inference. The pragmatic perspective, ed. by M. Papi and J. Verschueren, 61-129. Amsterdam: Benjamins.
Lewis, David. 1979. Scorekeeping in a language game. J. Philos. Logic 8. 339-59.
McCall, Storrs. 1963. Aristotle's modal syllogisms. Amsterdam: North–Holland.

Philosophy of Language Meets the Real World; or, When is "Enough" Enough?

Robin Lakoff
University of California, Berkeley

1.0. Introduction. Linguistic pragmatics, deriving as it does from two introspective fields, has a tendency to view its constructs and computations as determined in a vacuum, true for all time and all circumstances, independent of the context in which the utterance under investigation might be produced. This is the assumption, explicit or otherwise, in most discussion of the Cooperative Principle (CP) by Grice himself (1975) and others. Grice treats conversational logic as context-independent: what is "informative," how much and what kinds of computations speakers can assume hearers will make, and therefore how indirect someone can be, and still be understood as making an informative contribution, tend (not of course without exception) to be seen as universally constant, across cultural and discourse boundaries. That assumption enables us to propose and test our assumptions readily, constructing sentences and imagining their interpretations as we go along, allowing us to hypothesize and work with non-occurring infelicitous examples in order to determine the precise domain of our rule system. That introspective methodology is the inheritance pragmatics derives from its forebears, philosophy of language and transformational generative syntax. It has been of service to us, and will, in limits, continue to be. But it is time to look more closely at the idea of decontextualized pragmatics, to ask whether it can be, not only uninformative, but sometimes actually misleading.

Grice distinguishes among types of breakdowns of purely informative discourse, from violations of a Maxim so trivial as to go unnoticed:

(1)   A: I'm out of gas.
        B: There's a garage around the corner.

To cases of "flouting" that are noticeable, but computable by normal speakers:

(2)   A. Is Snarf qualified to be a Professor of Linguistics?
        B. He does produce exquisite footnotes.

But there is a world beyond flouting -- where even the most intrepid implicature cannot follow, and sense will not be made. Not only the Maxims, but the CP itself is abrogated. This is a large area, covering everything from the ravings of psychotics to a category which forms the basis of my discussion here: cases which, in one discourse genre with one set of discourse rules, would be treated as uninformative and unintelligible; but in another discourse context, will be subjected to special interpretive procedures and will thereby be made sense of. So the operations of the CP, like other pragmatic constructs, must be discussed relative to the context, whether cultural (cf. Keenan (1976) and Matsumoto (1989)) or, as here, textual.

1.1. The facts. The basis of my discussion is a legal case: a civil lawsuit for defamatory slander, in which I served as an expert witness for the plaintiff (Michael

McGarrigle, Persse. To appear. The influence of T. S. Eliot on Shakespeare. London: Lecky, Windrush and Bernstein. Cited in David Lodge, Small world. New York: Warner, 1984.

Mill, John Stuart. 1867. An examination of Sir William Hamilton's Philosophy, 3d ed. London: Longman.

Monck, W. H. S. 1881. Sir William Hamilton. London: Sampson, Low.

Morpurgo–Tagliabue, Guido. 1981. Grammar, logic and rhetoric in a pragmatic perspective. Possibilities and limitations of pragmatics, ed. by Herman Parret et al., 493-508. Amsterdam: Benjamins.

Nowell–Smith, P. H. 1954. Ethics. Harmondsworth: Pelican Books.

Oesterle, Jean T. (ed. & trans.) 1962. Aristotle: on interpretation. Commentary by Saint Thomas and Cajetan. Milwaukee: Marquette University Press.

O'Hair, S. G. 1969. Implications and meaning. Theoria 35: 38-54.

Rose, Lynn E. 1968. Aristotle's syllogistic. Springfield, Ill.: Charles C. Thomas.

Sapir, Edward. 1930. Totality. Language Monograph No. 6.

Sesmat, Augustin. 1951. Logique II. Paris: Hermann.

Sperber, Dan and Deirdre Wilson. 1986. Relevance. Cambridge: Harvard U. Press.

Stalnaker, Robert C. 1974. Pragmatic presuppositions. Semantics and philosophy, ed. by Milton Munitz & Peter Unger, 197-214. New York: New York University Press.

Strawson, P. F. 1950. On referring. Mind 59: 320-44.

Strawson, P. F. 1952. Introduction to logical theory. London: Methuen.

Suidas. Lexicographi Græci, ed. by I. Bekker. Berlin: Georg Reimer, 1854.

Thomason, Richmond H. 1987. Accommodation, meaning and implicature. Ms., University of Pittsburgh.

Tucci, Giuseppe. 1928. On the fragments from Dinnaga. JRAS 377-90.

Von Wright, G.H. 1951. An essay in modal logic. Amsterdam: North–Holland.

Washington, George. 1746/1988. George Washington's rules of civility & decent behavior in company and conversation. Reprinted, Boston: Applewood Books.

J. Weller, Edgar Morse, and Argentum Antiques, Ltd., d/b/a Argentum, vs. American Broadcasting Companies, Inc., KGO TV, Carol Ivy, Black Company, a corporation, White Company, a partnership, and Does One through Thirty: Marin County, California Superior Court #120451. The case will be referred to hereafter as Weller). The facts in the case are as follows:

Michael J. Weller and Edgar Morse, the plaintiffs, are the owners of an antique silver shop, Argentum Antiques, in San Francisco. During a six week period from February 23, 1984, to April 2, 1984, the KGO TV local news program broadcast a series of stories, each about 4 minutes in length, suggesting that Weller had: bought an antique candelabra which he knew the seller had acquired by fraud or theft; sold it at above market value to the DeYoung Museum; and lied to the DeYoung about the candelabra's provenance. After these broadcasts had begun to air, the shop's revenues dropped significantly. The loss in profits continued for several months after the last news report. As a result, the plaintiffs sued the TV station. The case came to trial in the spring of 1989.

After a trial of about six weeks' duration, a jury awarded the plaintiffs 2.3 million dollars in compensatory damages, with which, according to an article in the San Francisco Chronicle, they were "very pleased," and the defense counsel, "very disappointed." The case is currently under appeal.

1.2. The expert linguist's role. To prove defamatory slander, the plaintiffs had to prove to a jury that:

1. The statements made by KGO News were false; or rather, that the ordinary viewer was likely to derive an interpretation of the broadcasts that included false claims;

2. Those conclusions led to the plaintiffs suffering financial losses.

My job basically was to argue that the ordinary viewer's most probable interpretation went beyond the actual superficially accessible utterances themselves, and that normal linguistic inferential processes could and would allow viewers to draw conclusions that would create a bad opinion of the plaintiffs in the hearers' minds. I made no arguments about what the KGO News team intended to communicate, or meant, the illocutionary force (Austin (1962)), since there was no way it could be reliably determined. I assumed, rather, that I as in most respects an ordinary viewer, could use my own reactions as I watched the tape as a reasonably accurate representation of the utterances' perlocutionary force. The defense, trying to keep me from being qualified to testify, gave the judge as one of its arguments that, being an expert, I did not understand language as an ordinary speaker. But I would (and the plaintiff's lawyer did) argue that linguists process language more or less like everyone else -- or else we would either be in continual trouble, or very influential. Our training enables us to make our ordinary-speaker's inferential process explicit and subject it to analysis after the fact. But the inferences we draw are the same as those anyone else is apt to draw.

In general, linguists have trouble qualifying as expert witnesses. To clear their calendars, judges try to get through cases as quickly as possible. So their tendency is to try to weed out as much potential testimony as is legally permissible. One way to achieve this end is to be abstemious about qualifying new categories of experts. So doctors, psychologists, engineers, serologists and so forth, having qualified for years, are generally accepted without much fuss (though particular individuals can

be impeached, or their testimony shown to be inadmissible); but linguists, as a new category, have to be elaborately justified each time, and it's up to the lawyer who wants to use a linguist to prove that he or she should qualify, rather than being up to the judge to show why he or she should not. Linguistic testimony is frequently disallowed on the grounds that linguists do not have any knowledge not possessed by the average juror (since an expert witness' role is to provide information that only an expert knows): linguists' knowledge is only about how to talk, and every juror knows that. Sociolinguistic expertise is more readily admissible, since it concerns differential communicative practice from one group to another. Since the ordinary person would have access to, at best, only one of these systems, the expert could tell the jury something they do not know.

In <u>Weller,</u> the plaintiff's lawyer argued that my knowledge of language differed from that of the average juror in being explicit, and that therefore I could explicitly explain participants' inferential processes to jurors. Thus I qualified.

2.0. KGO's Narrative. I based my testimony on several viewings of videotapes of the series of broadcasts, as well as the examination of the written transcript of those broadcasts. Basically, the story conveyed by that series of broadcasts is as follows:

As the narrative opens, an antique appraiser, Jerry Durham, had been found guilty of insurance fraud. KGO's story mentioned that Durham had volunteered his services as an appraiser to the DeYoung Museum in San Francisco. After the report aired, an anonymous caller to KGO News linked Durham's name to that of Michael Weller, who had sold an antique silver candelabra to the DeYoung for $65,000. KGO referred to this transaction as a "sweetheart deal," with the price far too high for the condition of the candelabra. Weller claimed to have bought the piece from a Texas estate. According to reporter Carol Ivy, "sources in the community say that none of this is so, that the candelabra actually belonged to a sculptress named Barbara Herbert...and that pieces of her property, including the candlesticks, were...perhaps even stolen during the years just before her death." Notice the attribution of this crucial material to unnamed "sources."

Moreover, Ivy continues, "in another twist, [Durham] admits he once lived in an apartment in Barbara Herbert's house." Note the use of "admit," a verb which presupposes wrongdoing. Normally, living in someone's house is not wrong.

The plot, summarized in each program (with the logo "Museum Fraud?" at the top of the TV screen, and titled by the anchor, Van Amburg, as "the case of the confusing candelabra," is alleged to be as follows:
1. Jerry Durham, convicted felon, appropriated a pair of candelabra from Barbara Herbert.
2. Durham is an "associate" of Michael Weller, part-owner of Argentum Antiques.
3. Durham has served as an appraiser for the DeYoung.
4. Weller acquired the candelabra from Durham, knowing their dubious provenance.
5. Weller sold the candelabra to the DeYoung for $65,000.
6. That is an inflated price.

Of these claims, only #3 and #5 are uncontestably true; the others involve innuendo or the claims of anonymous sources. But those true claims are of no news interest, while the others are definitely newsworthy.

To argue for its claims, KGO relied largely on the on- and off-camera testimony of a former handyman of Barbara Herbert's, Lonnie Williams, an elderly

man who claimed to have polished Herbert's candelabra "nearly everyday for 30 years" and asserted on camera that the ones in the DeYoung were "definitely Mrs. Herbert's property, definitely, with no doubt, no doubt in my mind at all." But Herbert had died in 1979 and Williams' examination of the candlesticks took place in 1984. And there was no other corroboration of this claim. Moreover, an expert testified on the last program of the series that the provenance Weller claimed could well be correct (but couldn't be proved); and that the price paid by the museum was reasonable. At that, the KGO team went to considerable lengths to distance itself from its earlier claims:

> Carol Ivy: We must add that in previous stories we described a possible connection between another antique dealer, Jerry Durham, in the sale of the candelabra. Durham did once rent an apartment in the home of sculptress Barbara Herbert and Durham is acquainted with the man who made the sale to the DeYoung. However, we have no reason to suspect Mr. Durham is connected in any way with the candelabra....

which would seem to be a fair enough disclaimer, even if uttered rather late in the day. But Ivy continues at once:

> However, Van, there is still that unanswered interesting question which other museums tell me is an unusual situation and that is that we are not allowed to know who the last owner of these candelabra are [sic]. Now, we have Mr. Firestone's word for their authenticity and I think it's great that the DeYoung apparently got a good deal, but let's find out where they came from.

Van Amburg replies:

> Good question. Confusing candelabra. It will probably continue. Thank you, Carol.

And with that the story vanishes, never to surface again. So even with the disclaimer, the audience is left with a sense that there's more than meets the eye, the case is far from closed. KGO argues in its case that the disclaimer it made, plus its interview with the expert, Mr. Firestone, plus the fact that it hedged or otherwise downplayed the authority of most of its accusations, would lead a normal speaker to draw no negative conclusions about Weller or Argentum. Technically this may well be true. But in fact, I think normal speakers would interpret the speech events otherwise.

Let us grant KGO that, in a vacuum, references to "the case of the confusing candelabra" with its "twists," its "sources," its adversaries "refusing to speak to us," "not caring to be interviewed," "unwilling to tell us who in Texas owned the candelabra"; telling "their side of the story," "trying to talk to Lonnie Williams," and many similar examples; as opposed to their own side's "put[ting] together all the steps as you went along" [Amburg to Ivy]; "good memory" (Lonnie Williams', unsubstantiated); in which "efforts are being made" to get at the facts; they "hope the

museum relies on experts," because "I [Carol Ivy] have to rely on experts"; add up, logically, to nothing because they say nothing explicitly. But, under the special conditions of the news broadcast, the constant repetition of these claims, coupled with the daily repetition of the story itself, might well add up to something -- something which is ultimately reflected in the public's decision to stay away from Argentum in droves.

2.1. The frame and the scenario. In discussing the relation between the ostensibly bland surface communication, and the more controversial derived meaning, I will make use of two concepts, both in use (often in overlapping ways) in semantic and pragmatic theory over the last twenty years or so, though I will use them both in slightly idiosyncratic ways: the frame (cf. Minsky (1977) and Tannen (1979) and the scenario (cf. Fillmore (1975)). In this discussion, I will try to use frame consistently to refer to the set of assumptions based on the genre and function of a specific discourse, that all participants can be counted on to make to assign meaning to utterances within that discourse. The frame is part of one's normal expectations of how that kind of discourse generally works. (Examples here are the news broadcast and the courtroom.) I will describe by scenario the means devised intentionally by the active participants in nonreciprocal discourse such as news broadcasts to inculcate a particular desired attitude or state of mind in the passive participants: here, the mystery story or detective story and the we versus they or trust me scenario.

2.2. The news broadcast frame. Unlike ordinary conversation, people participate (as hearers) in television news broadcasts because they expect and hope to derive information from them. For that reason, ambiguous utterances or even utterances that in non-informative discourse types would be discounted or not made sense of will be interpreted as informative: efforts will be made to apply the rules of implicature even to utterances that otherwise are beyond the reach of CP entirely. Additionally, the purveyors of this information are assumed to be, by virtue of their position, authoritative persons (they must know what they're talking about, or they wouldn't have those jobs). What we might dismiss as mere gossip or guesswork in ordinary conversation becomes "information," that is, fact, in a TV news setting. This frame is put in place by, first of all, the fact of the viewer's tuning in; then by the familiar accoutrements of the news program: the anchor at his desk, the reporter with her microphone, the identifying logo at the top of the screen, the serious facial expressions of the news team while it is being "newsy," the measured and non-conversational delivery of the script. (I am informed that everything said by anchors and reporters on local news shows is pre-scripted, even the ad libs.)

2.3. The mystery and investigative reporter scenarios. In this series of broadcasts, the informational nature of the frame is emphasized by the institution of the mystery or detective story and investigative reporter scenarios. Both of these involve language or pictures (often both) that invoke these familiar fictional genres (and mostly serve little other informative purpose). Like the news frame, these scenarios are information-based: the detective and the reporter are there to ferret out the facts. But they develop the frame further. The news format makes no assumption of obstructionism, and therefore no claim that there are reasons why the facts are not clear. But in both of these genres, the protagonist is fighting against odds to get the facts; there is wrongdoing, the miscreants are seeking to hide the evidence; if it is

not uncovered and the bad guys brought to justice, <u>we</u> (protagonist and people of good will everywhere) will suffer. These scenarios provide a bridge between the staid and distant communication of the <u>news</u> frame, and the involved and mutual interaction implied in the <u>trust me</u> scenario.

Van Amburg refers on several occasions to the "case of the confusing candelabra," a title unavoidably reminiscent of a classic detective series, Perry Mason. At other times he shortens it to refer to "confusing candelabra," and often comments on how "confusing" things are: all of these recalling the Mason reference. He often calls the situation a "mystery" or "mysterious," and refers to "strange" or "bizarre twists." Ivy often concurs in these assessments. At times the very fuzziness, rather than denying the existence of meaning, lends an air of euphemistic avoidance: "Yes, there is meaning here, but it's so awful we can only hint at it":

> Van Amburg: Yes, it's still a mystery. I'm not trying to put you [Ivy] on the spot and make a decision, but it's always "buyer beware." But we've heard about cases like this in the past where there have [sic] been an awful lot of money spent on one end and then it would be written off at the other end and all kinds of things would happen in between.

> Carol Ivy: And we don't know enough about it yet actually to make those kinds of allegations and those kinds of allegations may never be made. The question is were those Barbara Herbert's candlesticks and are they worth the price the DeYoung paid or are they in fact from an estate in Texas and are they the real Duke of Cumberland candlesticks which the DeYoung thinks they are?

> Van Amburg: As they used to say in radio: "stay tuned."

> Carol Ivy: "Stay tuned."

On its face, this discourse looks absolutely nonsensical, almost childish with its strings of coordinate conjunctions and repetitions of words and phrases. Amburg's phrasing is totally empty semantically: what is the "one end," and "the other," what are "cases like this," what are "all kinds of things" and "in between"? Ivy, for her part, denies the possibility of knowing, but her use of "yet" covertly injects the presupposition that there is knowledge that will eventually come out. She frames most of her remark as a question, theoretically without informative content; but note that the question is multiply compound. And in each of the choices (are they Barbara Herbert's <u>or</u> not? Are they worth their price <u>or</u> not?) one possibility has no news value, and therefore in this context violates the Maxim of Quantity. The other, the more interesting, is informative and suggestive of wrongdoing. Likewise, Ivy's rhetorical compound question at the top of the first broadcast: "Did the DeYoung Museum pay a grossly inflated price or did they actually acquire an excellent addition to their collection?" So, if this is a <u>news broadcast</u>, if indeed we are operating within the domain of a <u>mystery scenario</u>, the informative and accusatory choice is the only one that makes sense.

Ivy's reports stress the fact that the adversaries intentionally violate the Cooperative Principle; but she violates it in turn by not being explicit about their uncooperativeness, or making it clear that that's the intention of her lexical choices. But these choices are far too numerous to be due to chance: as noted earlier, Weller and the DeYoung "refuse" to give her information (rather than just not giving it); do "not care to be interviewed"; we are "continuing to ask to photograph the candelabra," suggesting foot-dragging by the DeYoung; Weller is "reportedly" out of town. Overall, they are lying, hiding; we must find out the truth.

Alongside the mystery scenario is the investigative reporter scenario. Like the detective, the reporter must go against opposition to get the facts, in order to protect society (the audience). This overlay shows up mostly in pictures: Carol Ivy striding down the mean streets (past a store called "Video Conspiracy") in a trench coat; Carol Ivy talking to "sources" over the telephone as the voiceover summarizes the conversation. Neither of these pictures in any way furthers the informative content of the story, any more than do the mystery details. The repeated references to "sources," too, is a way of invoking this scenario, even as it dilutes the assertional force of the speech acts involved by refusing to identify a responsible speaker (rather like the agentless passive). We might argue that they interfere with intelligibility by violating Quantity, Relevance and Manner (a grouch might add Quality).

2.4. We versus they. These scenarios create a bridge linking the apparent impersonal objectivity of the news frame with the creation of trust between viewers and the news team. The problem with purely objective reports, from the broadcaster's perspective, is that viewers might refuse to give them credence, or make the extra effort to override the obfuscations. But if the audience is made an honorary member of the team, one of us; if it is clear that the news team is doing its uncovering job for the audience's sake; if the point is made that the activities described will hurt you, the public; and if the subjects of the investigation are turned into a group of hostile they whom we must fight together, trust will allow hearers to assume speakers are making sense even when, technically, they aren't; to override floutings and derive clear and unambiguous meaning. It is not unconnected with Bernstein's (1973) notion of elaborated and restricted code: the better we know and like someone, the less clear they have to be for us to be able to understand; and the more fuzzy someone is, once they have won our hearts, the more we trust them because the fuzz shows they trust us, we all speak the same language. So the news frame encourages us to understand apparent unclarity as meaningful, because the discourse genre itself is defined as informative; the we/they scenario encourages us to make sense of confusion because it's our kind of confusion, they wouldn't be imprecise if they didn't believe we wanted to make the effort to understand. Additionally, once our trust is won, its winners achieve still more authority in our eyes: they are our people, they must know what they're talking about.

A major force in the creation of the we is the "happy talk" that is, these days, an essential part of local news broadcasts. Pioneered perhaps twenty years ago, "happy talk" is the inconsequential patter between anchors and reporters, talk that does nothing to bolster the information content, but is purely interactive. You can tell it's happening when members of the team make eye contact with each other or call one another by name; when they laugh; when they comment to one another about the story (which at the same time models to the audience how they are to

react):

> Amburg: Really interesting story. You put together all of the steps
> as it went along. Uh, what do you think?

> Ivy: I think it's still a mystery at this point.

Of course, this also reinforces the <u>news</u> frame and the <u>mystery</u> and <u>investigative</u> <u>reporter</u> scenarios. But in a format where time is precious and information paramount, valuable moments cannot be spent on sheer emptiness: this chatter must be doing something. It is: it's showing us that Van and Carol are regular guys who react like you and me, are interested as you and I should be; and therefore are to be trusted and believed.

Then the news team comments on their sharing with the audience a desire to root out evil everywhere. Not only are the news team a team; the audience is also part of it. The references to "sources in the community" begin the process. Carol Ivy addresses her hearers directly: "Yesterday I told you...I showed you...." "I look forward to seeing whatever proof they have of the history and I am certain so will the many viewers who have expressed interest in the story, particularly since the DeYoung is a public museum that many of them visit frequently." The DeYoung, Weller, Argentum, and Durham are lumped together as an underhanded, unreliable, and uncooperative <u>they</u>: the enemy, those with something to hide.

At trial, the defense argued that the transcript contained many caveats and retractions, forms essentially subverting or denying the story's informative content. Especially toward the end, after Weller's attorneys contacted KGO, they began to hedge anything and everything. But often they retract with one hand and insert with another: "We don't know enough about it <u>yet</u>"; frequent use of <u>may</u>, as well as <u>sources say</u>, virtually a speech act hedge; "Again, we are not alleging anything at this point. We are simply saying that there are two stories about these candelabra." True, officially these create a flouting of Quantity: "We have nothing to report." But if you have nothing to report, you don't put it on the six o'clock news. So, if anything is to be made of the disclaimers, they deepen the air of mystery and romance: <u>something</u> is going on, hints are being bruited about. Since <u>we</u> are in this together, it is <u>our</u> job (the audience's) to figure out what our friends mean, thus helping nab the baddies.

3.0. The Trial Frame. And this brings us, finally, to the outer envelope of this paper, the trial and the expert's role in it. Like the <u>news</u> frame, the <u>trial</u> frame is cast as maximally informative: we are here to get at <u>the truth</u>, to discern <u>the facts.</u> Like the <u>investigative reporter</u> scenario, the discourse involves an adversary seeking to protect himself by keeping the facts from coming out. The witness is the conduit through which facts are to surface. Indeed, the witness' role is made clear before he or she even assumes the stand: the oath requires one to tell the truth, the whole truth, and nothing but the truth. This can be read as a strict requirement of adherence to the Maxims. But in fact (as the CP also tells us) if there were not some communicatively salient reason to require the explicit exaction of an oath, no oath would be taken: it would violate Quantity. So the insistence on the oath means that it would be interesting indeed if the witness were to adhere completely to its tenets. Thus, the job of the cross-examination: to find discrepancies between the witness'

testimony and his or her promise to be fully informative.

Expert witnesses, as opposed to eyewitnesses, have something of a superior position in a courtroom: they know something the jurors don't, and might find hard to understand. Thus they are authoritative by virtue of their role alone, unlike eyewitnesses who must rely more on personal style to be persuasive (Lind and O'Barr (1979)). But experts are still cautioned: don't be pompous, don't ramble, try to be clear. The plaintiff's lawyer had set up a blackboard. He threw at me various technical terms of the trade (illocutionary force, hedge, Maxim of Quantity, implicature, modal auxiliary, frame, scenario...). For the first part of the direct examination I wrote these on the board, defined them, provided examples. The judge remarked: "The final exam will take place this afternoon." We then went through the KGO transcripts, with me commenting on everything that might be relevant, prompted by questions. It was pretty straightforward, even if it felt like cramming the contents of a semester's pragmatics course into two hours.

It is in the cross-examination that the game is played. The witness is juggling several sorts of discourse roles and expectations. The opposing lawyer is relying on a deposition that, in this case, had been made almost three years before; I had reviewed it, but it was about two hundred pages long. If I contradicted my deposed testimony, I would be in trouble. I had also been warned to stick to the point: don't expatiate on answers, not only because you'll lose the jury, but because you may "open the door": permit a line of questioning you didn't want to get into. So the courtroom requirement of strict informativeness on the one hand seemed reinforced by these constraints (one was not encouraged to digress); but here too, the notion that strict adherence to the Maxims is the clearest or most informative way to talk is truer in the abstract than in actuality. In ordinary conversation especially, we rely on backgrounding, on extraneous but useful filling-in, on partial disclaimers, to help us make sense of the "gist." In ordinary conversation, if your interlocutor presents a position diametrically opposite yours, you may try to split the difference to maximize rapport: "Ye-es, that's right, but still...." In cross examination, such an admission, while cooperative, could be deadly. At first, worried about being shredded by my questioner, I answered, unhappily, "yes" or "no." But I soon learned that, if I just prefaced my "unresponsive" answer with "We-el yes, but...." I could get away with a comfortable deviation from direct informativeness and be <u>really</u> informative. In response to a question, Doesn't the use of a question dilute a female speaker's authority?, the most informative answer was not a mere "yes" or "no," but a response that considered the influence of discourse context and the role of the speaker. Yes, I had said in earlier work (which the lawyer was waving in the air like a Bible) that women lost authority when they phrased assertions as questions, but no, Carol Ivy's clout was not diminished thereby here. For one thing, her questions were rhetorical. For another, she came into the discourse with the authority of her role.

Then, too, a witness is enjoined to be explicit: to answer questions orally, no nods, no backchannels. But this advice is hard to follow when one is asked obvious or repetitive questions, to which in ordinary conversation one would barely nod or grunt: to do more would be too informative, in the sense of suggesting that one was supplying real information, and would thus ultimately violate Quality as deceptive. But in court, the nod doesn't get into the reporter's transcript, and one is

reprimanded by attorney or judge, and loses face before the jury.

Courtroom discourse is special in that it is adversarial. While most forms of discourse (ordinary conversation, TV news, classroom, and therapy, for instance) involve participants at least theoretically united in their desire to share information (only theoretically or partially, for most of these, to be sure), in court information that is advantageous to one party is apt to be a liability to the other, so for all witnesses, there is a balance between adherence to the Maxims, saying everything that is useful, true, and relevant, in a clear way, and thereby doing one's sworn job; and risking damage of one sort or another.

As with a broadcast, the flow of information in court is both presumed to be especially informative, but for extrinsic reasons often very much not so. The Cooperative Principle helps us as investigators and as participants to navigate these treacherous waters; but only if we are willing to build special contextual assumptions into the system.

References

Austin, J. L. 1962. How to do things with words. Oxford: Clarendon Press.

Bernstein, Basil. 1973. Class, codes, and control, vol. 1. St. Albans, Herts.: Paladin.

Fillmore, Charles. 1975. An alternative to checklist theories of meaning. Proceedings of the first meeting of the Berkeley Linguistic Society, pp. 123-32.

Gifis, Steven H. 1984. Law Dictionary, Second Edition. Woodbury, N. Y.: Barron's Educational Series.

Grice, H. P. 1975. Logic and conversation. Syntax and semantics 3: Speech acts, ed. by Peter Cole and Jerry L. Morgan. New York: Academic Press, pp. 41-58.

Lind, E. Allen and William M. O'Barr. 1979. The social significance of speech in the courtroom. Language and social psychology, ed. by Howard Giles and Robert St. Clair. Oxford: Basil Blackwell, pp. 66-87.

Matsumoto, Yoshiko. 1989. Politeness and conversational universals -- observations from Japanese. Multilingua 8-2/3, pp. 223-48.

Minsky, Marvin. 1977. Frame-system theory. Thinking: Readings in cognitive science, ed. by P. N. Laird and P. Johnson. Cambridge: Cambridge University Press.

Keenan, Elinor O. 1976. The universality of conversational implicature. Language in society 5, pp. 67-80.

Tannen, Deborah. 1979. What's in a frame? Surface evidence for underlying expectations. New directions in discourse processing, volume 2, ed. by Roy O. Freedle. Norwood, N. J.: Ablex.

Syntax and Discourse: A Look at Resumptive Pronouns*
Ellen F. Prince
University of Pennsylvania

1. Introduction.
Whenever a grammar has two or more forms that are in free variation from the point of view of truth-conditional semantics, it is worthwhile investigating whether there are in fact pragmatic differences that distinguish them. In this paper, I shall investigate the discourse functions of a particular syntactic form in Yiddish and its correlate in English, relative clauses containing a resumptive pronoun, comparing them with gap-containing relative clauses in each language.

Standard Yiddish has, alongside of gap-containing relative clauses, as in 1, relative clauses containing so-called resumptive pronouns, as in 2:

(1) a. ...a yid vos [e] hot geheysn rabinovitsh... (RP81)
...a Jew that [e] has be-called Rabinowitz...
'...a guy who [e] was named Rabinowitz...'
b. ...an alte yidene vos [e] iz nebekh gelegn in bet (RP230)
...an old Jewess who [e] is PRT lain in bed...
'...an old bag who [e] was lying in bed, the poor thing...'
c. ...di mayse vos ikh vel [e] aykh dertseyln... (RP 63)
...the story that I will [e] you tell...
'...the story that I'll tell you [e]...'

(2) a. ...a yidene vos zi hot geheysn yente... (RP29)
...a Jewess that she has been-called Yenta...
'...an old bag that [she] was named Yenta...'
b. ...a yid vos er iz geven a groyser lamdn un a gvir... (RP178)
...a Jew that he is been a big scholar and a rich-man...
'...a guy who [he] was a big scholar and a rich man...'
c. ...mentshn vos a shlang hot zey gebisn... (RP248)
...people that a snake has them bitten...
'...people that a snake has bitten [them]...'

The situation in English is somewhat different: relative clauses with resumptive pronouns are officially ungrammatical in English (Langendoen 1970, Kroch 1981, Sells 1987, inter alia; but cf. Kayne 1981). However, they are in fact not uncommon in speech. Some examples are given in 3 and 4:

(3) a. There are always **guests who I am curious about what *they* are going to say.** (AK:Dick Cavett)
b. **The only one we could see *her* figure** was Number 2. (AK:Kitty Carlisle, To Tell the Truth, 9/8/81)
c. Let's get to **our first guest, who I asked for and was so delighted that *he* could make it.** (AK:Orson Wells, Tonight Show)

(4) a. They were just towed across the Midway onto **the bridle path, where they were just sitting *there* peacefully.** (AK:Laurence Horn)
b. That's a suggestion of yours which I followed, **which I didn't even want to do *that*.** (AK:Gregory Ward)
c. I have **a friend who *she* does all the platters.** (AK:Ellen Prince)

Thus, we may pose the following questions: What prompts a Yiddish or English speaker to utter a gap-containing vs. a pronoun-containing relative clause in a particular context? Is

the distribution of the two types of clauses random, or does one type do some work that the other does not do, resulting in a nonrandom distribution? Finally, in the event that both the Yiddish and the English resumptive pronoun clauses have some function, is it the same function? I shall now present findings from an ongoing study of the functions of resumptive pronoun clauses in Yiddish and English discourse.

2. Resumptive pronoun clauses and island violations: a processing function.

2.1. English.

What mention there has been in the literature of English resumptive pronouns has generally taken the approach that they are ungrammatical but are ways of salvaging a sentence that a speaker has started without realizing that it is impossible or at least difficult to finish it grammatically, 'making the best of a bad job,' as Langendoen puts it (1970:104). See also Bever, Carroll, and Hurtig 1976 and Kroch 1981. Thus we find many resumptive pronouns in island environments, e.g. indirect questions as in 3a, left branches as in 3b, and following a complementizer in embedded sentences, as in 3c. As would be expected, attempted extraction from other island environments may also result in resumptive pronouns, e.g. relative clauses as in 5a, adverbial clauses as in 5b, and coordinate structures, as in 5c:

(5) a. **That asshole X, who I loathe and despise the ground *he* walks on,** pointed out that... (AK: William Labov)
    b. Apparently there are such things as **bees in the area which if you're stung by *them*, you die.** (AK:D.R.)
    c. **...or those eighteen upon whom the tower in Siloam fell, and slew *them*.** (AK: Luke 13.4)

Furthermore, we find such 'breakdown' resumptive pronouns in environments where extraction is possible but difficult, either because some material has been preposed, adding to the distance (and time) between the head and the would-be gap, as in 6a, or in the case of a stacked relative, presumably for the same reason of distance/time, as in 6b, or in combinations of the above, as in 6c, 6d:

(6) a. **This is the one if you tell her to go away *she* stays, if you tell her to stay, *she* leaves.** (AK: M.K.)
    b. **There are two plants that need more light, that we have a good chance of getting *them* to bloom if we give them more light.** (AK:Anthony Kroch)
    c. **I got a new pen the other day that someone gave me that if you press the clip *it* shows the time.** (AK:J.C.)
    d. **What does it print out some statement that's really obscure that the average reader is going to look at *it* and wonder what *it* means.** (AK:T.X.)

Finally, it is possible that some extractions fail because the constituent to be extracted represents an argument low on the Accessibility Hierarchy (Keenan and Comrie 1977) and a resumptive pronoun occurs. Perhaps the datives in 7 are such cases:

(7) a. **...the man who this made *him* feel sad...** (AK:P/S)
    b. **Some of the same judges who we told *them* that if you mess with John Africa, you going to...** (AK: MOVE member)
    c. **He looks like one of those guys you got to be careful throwing *them* fastballs.** (AK:Tim McCarver, KYW Radio)

2.2. Yiddish.

The situation in Yiddish is very similar in that resumptive pronouns occur when extraction would violate an island constraint or when processing is difficult for some other reason. However, the Yiddish corpus used for the present study is a written one and understandably has few such instances. All the same, we do find one possessive (8a), one embedded subject (8b) and a number of datives, e.g. 8c:

(8) a. dos iz oykh a shlekhte krenk,
         this is also a bad sickness,
         di gelt vos *zeyer* har hot fun zey tsores. (RP193)
         the riches that *their* owner has from them troubles.
     b. bay eynem a keyser iz geven a minister mit **a sheyner bord**
         by one a czar is been a minister with **a beautiful beard**
         vos er flegt zogn az *zayn bord* iz di shenste fun der velt.
         that he used say that *his beard* is the most-beautiful of the world.
         'This one czar had a minister with **a beautiful beard that he used to say [*his beard*] was the most beautiful in the world.**'
     c. es iz geven amol **a melamed, vos es iz *im* zeyer shlekht gegangen.**
         it is been once **a teacher that it is *him* very bad gone.**
         'There was once **a teacher who [*he*] was very badly off.**' (RP184)

In fact, I suspect that certain datives are simply nonextractable; these are the datives in 'subjectless' sentences, of which 8c is an example. That is, datives in ordinary subject-containing sentences can be extracted so long as the complementizer is one that can be inflected, as in 9b. However, an informant found it nearly impossible to extract a dative in the subjectless sentence corresponding to 8c, as in 10a; she was unable to produce 9b and, when I produced it, could say only 'Well, I guess that's grammatical,' a response motivated perhaps more by her fine logic than by her equally fine linguistic intuitions. Furthermore, it was even harder for her to extract the dative from subjectless sentences with *zayn* 'be' as their main verb, as in 10b, and she accepted easily only the resumptive pronoun forms 10c and 10d:

(9) a. ...eyn eyntsikn zun, vos **im** gehert di gantse yerushe... (RP167)
         ...one single son, that **him** belongs the whole inheritance...
         '...an only son, to whom the whole inheritance belongs [**to him**]...'
     b. ...eyn eyntsikn zun, vemen/velkhn es gehert di gantse yerushe
         ...one single son whom/whom it belongs the whole inheritance

(10) a. ??...a melamed vemen/velkhn s'iz [e] zeyer shlekht gegangen.
          ...a teacher whom/whom it is [e] very bad gone
     b   ?*...a mentsh vemen/velkhn s'iz [e] kalt
          ...a person whom/whom it is [e] cold
          '...a person who [e] is=feels cold'
     c. ...a mentsh vos s'iz **im** kalt
          ...a person that it is **him** cold
     d   ...a mentsh vos **im** iz kalt
          ...a person that **him** is cold

Thus we see that the processing function attributed to English resumptive pronoun clauses in the literature, that is, the function of permitting speakers to finish sentences which have been poorly planned from the point of view of the grammar, may also be ascribed to Yiddish resumptive pronoun clauses.

485

2.3. Where the processing explanation fails.

One looks for discourse functions of a particular syntactic form when it is clear that speakers have options; in the case of resumptive pronoun clauses with processing functions, there are no options and hence no reason to look for discourse explanations. The fact remains, however, that both Yiddish and English speakers produce many resumptive pronoun clauses which do not fall into the processing-function category in any obvious way, i.e. where no syntactic salvaging appears to be needed. Consider, for example, 2 and 4 above. Such cases, where the relative clause consists of a simple sentence with the subject or object relativized and with no intervening material and no obvious sign of dysfluency, are cases where we have to assume that the resumptive pronoun clauses are grammatical alternatives to gap-containing clauses, and it is here that we shall try to see whether they do any discourse work for the speaker.

3. Previous claims about the discourse functions of resumptive pronouns.

While there has been no discussion in the literature of Yiddish resumptive pronouns and no specific claims about non-processing functions for English resumptive pronouns, some claims have been made about the discourse function of resumptive pronouns in other languages, which we shall now consider.

3.1. The 'concept' claim.

First, resumptive pronouns in island-type environments have been noted in a number of languages in addition to English, e.g. the Scandinavian languages (Engdahl 1979, Maling and Zaenen 1980). Accounts differ as to whether they are grammatical or not, but their occurrence and distribution has been recognized.

    In addition, resumptive pronouns in simple contexts have been noted in other languages, in particular Hebrew and the Celtic languages. With respect to Hebrew, Doron 1982 claims that the antecedent of a resumptive pronoun has widest scope in the clause that contains the pronoun, whereas the antecedent of a gap does not necessarily. For example, Doron contrasts 11a with 11b; while the gap-containing 11a can have a referential or non-referential understanding for 'the woman', the pronoun-containing 11b can have only the referential reading, paraphrased in 11c:

(11) a. dani yimca et ha-iSa Se-hu mexapes *[e]*. [= Doron 1982, ex. 49]
        Dani will-find ACC the-woman that-he seeks *[e]*
    b. dani yimca et ha-iSa Se-hu mexapes *ota*. [= Doron 1982, ex. 50]
        Dani will-find ACC the-woman that-he seeks *her*
    c. There is a woman that Dani is seeking and he will find this woman.

Sells 1987 concurs, arguing that resumptive pronouns cannot have a 'concept' interpretation, his term for the nonspecific/de dicto/narrow scope reading possible in 11a but not 11b. Whether his claim is restricted to Hebrew or is generalizable universally, as he implicates, is not clear.

    In any event, Erteschik-Shir 1989 disagrees with the claim even for Hebrew and finds a context in which Sells' own example can have a 'concept' interpretation. Ariel 1990 likewise disagrees with Sells' claim for Hebrew, providing counterexamples like 12:

(12) higia ha-zman she-navin she-pashut eyn
      arrived the-time that-we-should-understand that-simply there-is-not
      be-nimca ha-muamad ha-ideali she-*oto* anaxnu mexapsim.
      in-existence the-candidate the-ideal that-*him* we are-seeking
      'It is about time that we realized that **the ideal candidate that** *[him]* **we are looking for simply does not exist.**' [= Ariel 1990, ex. 26]

Regardless of the situation in Hebrew, Sells' and Doron's claim of necessary referentiality clearly does not hold for Yiddish. The Yiddish equivalent of 11b is 13:

(13)   dani vet trefn **di froy vos er zukht** *zi*.
       Dani will find **the woman that he seeks** *her*

The resumptive pronoun clause in 13 is perfectly grammatical with the reading in 11c, although it is not particularly felicitous for reasons which will become clear later. In the corpus, we find an example of a resumptive pronoun clause where the head is nonreferential, shown in 14a, and one from a letter from a Yiddish speaker in Bucharest is shown in 14b:

(14)   a.   [A messenger comes to tell a Jewish community that a priest wishes to have a
            Bible contest with the Jews.]
            di shtetl zol shteln **a yidn vos** *er* **zol geyn oysfregn dem galekh**,
            the village shall place **a Jew that** *he* **shall interrogate the priest,**
            un der galekh zol oysfregn dem yidn. (RP:220)
            and the priest shall interrogate the Jew.
            'The village should pick **a Jew who** *[he]* **will question the priest...**'
       b.   er hot nokh nisht gefunen **dos meydl vos er zol derfiln az er hot** *zi* **lib**.
            he has still not found **the girl that he shall feel that he has** *her* **love**
            'He still hasn't found **the girl he feels he loves** *[her]*.' (Y. G.)

Both 14a and 14b clearly involve 'concept' interpretations. The resumptive pronoun in 14b may well have a processing motivation, since the extraction site is in a complement, but neither Doron nor Sells make any distinction between island-type cases and simple cases, and therefore 14a,b show that their claim that resumptive pronouns may not refer to nonspecific heads does not apply to Yiddish. We shall return to this matter below.

Likewise, it is not the case that 'concept' interpretations are impossible in English. A number are found in the corpus, e.g. those in 15:

(15)   a.   ...in case there's **important plans that a robber wants to steal**
            them. (AK:D.K.)
       b.   If there is **any message that she can forward** *it* **to us**, then...
            (AK:William Labov)
       c.   I don't think there's **a company in the world that the**
            stockholders would allow a company to copy *it*. (AK:engineer
            on TV)
       d.   You get **a rack that the bike will sit on** *it*. (AK: E.K.)

3.2. The 'restrictive-set' claim.
Second, Erteschik-Shir (1989) argues that Hebrew resumptive pronoun clauses are used when a 'restrictive-set' reading is intended. She contrasts 16a, where a gap is preferred, with 16b, where a pronoun is preferred:

(16)   a.   [In the context where H knows that S bought a dress.]
            hine **ha-simla she-kaniti** *[e]*. [= Erteschik-Shir 1989, ex. 12]
            here-is **the-dress that-I-bought** *[e]*
       b.   [In the context where H knows S had three dresses in mind originally.]
            hine **ha-simla she-kaniti** *ota*.
            here-is **the-dress that-I-bought** *it*.

However, the notion of 'restrictive-set' does not seem relevant for English or Yiddish resumptive pronouns. While the Yiddish equivalent of 16b is clearly grammatical, it is not

felicitous, certainly not in the restrictive-set understanding forced by the context. Furthermore, none of the naturally-occurring examples in the corpus have a restrictive-set reading. As for English, while I am reluctant to pass judgment on the grammaticality of the English equivalent of 16b, it is likewise the case that none of the naturally-occurring examples in the corpus are contrastive in this way.

3.3. The 'accessibility' claim.
In her discussion of the cognitive 'accessibility' of discourse entities and their linguistic representations, Ariel 1990 argues for a pragmatic explanation for the ungrammaticality of relative-clause-initial resumptive pronoun subjects in Hebrew. Her argument is that the antecedent in such cases is maximally 'accessible', being immediately prior to the extraction site, and that resumptive pronouns occur when antecedents are low on the accessibility scale. Whatever the correctness of such an account for Hebrew, it certainly cannot be maintained for Yiddish or English, since the majority of naturally-occurring resumptive pronouns in both the Yiddish and the English corpuses are initial subjects. See, for example, 2a,b and 14a above for Yiddish, 4c above for English.

4. The present study.
We shall now turn to the present study of how resumptive pronoun relative clauses are used in Yiddish and English. First, however, a description of the corpuses on which the study was based is in order.

4.1. The corpuses.
4.1.1. The Yiddish corpus.
The corpus for the Yiddish study is taken from *Royte pomerantsn*, a 200-page compilation of anecdotes gathered and edited by Immanuel Olsvanger and published in a romanized version by Schocken Press (1947). This volume contains 169 headed relatives, of which 31/18% have resumptive pronouns and 138/82% have gaps. Of the 31 resumptive pronoun clauses, three were omitted from the study, since their presence could be explained by the processing explanation given above and since they have no gapped correlate. (These three are shown in 8 above.) Other tokens which possibly have a processing explanation but which do have a gapped correlate, e.g. 9a, were included. The rationale for this was that these tokens are from a published work where the author had ample time to plan and edit, where he had a choice and chose a resumptive pronoun, and it is exactly that choice that we are investigating. Thus we are left with 166 headed relatives where the author had a choice of clause-type, choosing a gap in 138/83% of them, a resumptive pronoun in the remaining 28/17%.

4.1.2. The English corpus.
The English study is based on two very different types of data, different both from each other and from the Yiddish data. The English resumptive pronoun clauses are from a collection of 539 naturally-occurring resumptive pronoun clauses with their immediate contexts collected about 9 years ago by Anthony Kroch. These tokens are taken from a wide variety of situations, from casual conversation to classroom discourse to radio and TV newscasts to the New Testament, and the speakers represent an equally wide range, from inner city adolescents to plumbers to Philadelphia aristocrats to Luke. For the present study, only those speakers who were known to be nonnative speakers of American or British English were removed from the corpus. Although no other tokens were omitted because of their speaker, about 70% were in fact removed since it was felt that they had a sufficiently strong processing explanation. That is, all cases where the resumptive pronoun is in an island, e.g. 3 and 5 above, and where there is preposed material intervening between the head and the extraction site, e.g. 6a, or where the extraction site is in a non-initial stacked relative, e.g. 6b, were omitted from the study. For the sake of consistency, datives were not omitted, since they had not been omitted from the Yiddish

corpus, although the English corpus, being spontaneously produced, is more likely to contain breakdowns than the Yiddish corpus. After all these omissions, 158 resumptive pronoun clauses were left where no processing explanation is obvious and where some other motivation for their having been uttered seems likely.

In order to compare these resumptive pronoun clauses with gap-containing clauses, we used a second corpus, the 115 gap-containing clauses occurring in a taped career-counseling session, the transcript of which is on-line in the Computer and Information Science Department at the University of Pennsylvania. The participants are the counselor and a college-age woman, both native speakers of an unmarked variety of American English. [1]

4.1.3. Totals.
The figures for the two corpuses and their provenance are presented in 17:

(17)

	Yiddish		English	
	Gap	Pronoun	Gap	Pronoun
Source:	*Royte pomerantsn.*	*Royte pomerantsn*	counseling	A. Kroch
Total:	138	31	115	539
Foreign:	-	-	-	20/4%
Island+:	-	3/10%	-	361/67%
Corpus:	138/100%	28/90%	115/100%	158/29%

4.2. Discourse functions of resumptive pronouns: distributional facts.
4.2.1. Yiddish.
When we compare gap-containing with pronoun-containing relative clauses in Yiddish, a number of patterns emerge. First, we notice that the head of the resumptive pronoun clauses has a significantly greater tendency to be indefinite than does the head of gap-containing clauses, as shown in 18:

(18) **Yiddish:**

	Gap-containing clauses	Resumptive pronoun clauses
Def. head:	60	3
Indef. head:	76	25
Adj. head:	2	-

Chi-square = 10.951   p < 0.001

Second, when we look at the distribution of these definite and indefinite heads, we see that the distinction is relevant only in restrictive relative clauses, as shown in 19:

(19) **Yiddish:**

	Nonrestrictive		Restrictive	
	Gap	Pronoun	Gap	Pronoun
Definite head:	6	3	54	-
Indefinite head:	3	1	73	24

Chi-square = 0.090   Chi-square = 15.886
p < 0.80 [N.S.]   p < 0.001

Furthermore, if we look at the distribution of definite and indefinite heads in restrictive and nonrestrictive clauses only with gaps, we see that it is not significant, as shown in 20:

(20) **Yiddish:**

	Gap-containing relative clauses only	
	Nonrestrictive	Restrictive
Definite head:	6	54
Indefinite head:	3	73

Chi-square = 1.988   p < 0.20 [N.S.]

Not surprisingly, perhaps, from these figures, if we look at the distribution of definite and indefinite in restrictive and nonrestrictive clauses only with resumptive pronouns, we see that it is highly significant, shown in 21:

(21) **Yiddish:**

	Resumptive pronoun relative clauses only	
	Nonrestrictive	Restrictive
Definite head:	3	0
Indefinite head	1	24

Chi-square = 20.160   p < 0.001

Although the numbers are small, we see a clear pattern in that resumptive pronoun relative clauses in Yiddish strongly favor either nonrestrictives or else restrictives with indefinite heads, whereas no such favoring appears in gap-containing relative clauses. And, in fact, when I constructed resumptive pronoun relative clauses that were restrictive and definite, informants rejected them, e.g. 22:

(22) a. [Context: There were two guys, one who had a pass and one who didn't.]
#makht **der yid vos** *er* **hot gehat a pas,** 'zorg zikh nit.'
makes **the Jew that** *he* **has had a pass,** 'worry REFL not'
'So the guy who *[he]* had a pass said, 'Don't worry."
b. #di froy hot farloym   **vos zi hot** *es* **gehat.**
the woman has lost all **that she has** *it* **had.**
c. #lomir prubim a   ton dem ershtn yidn **vos mir veln** *im* **trefn.**
let-us try an ask do **the first Jew that we will** *him* **find**
'Let's ask **the first guy we meet** *[him]*.'

4.2.2. English.

We shall now turn to the distribution facts for the English corpus. First, the distribution of definites/indefinites in gap-containing vs. resumptive pronoun clauses is not significant, shown in 23:

(23) **English:**

	Gap-containing clauses	Resumptive pronoun clauses
Def. head:	35	56
Indef. head:	80	102

Chi-square: 0.751   p < 0.50 [N.S.]

However, if we separate restrictive and nonrestrictive clauses, the situation looks more like what we have already seen in Yiddish, as shown in 24:

(24) **English:**

	Nonrestrictive		Restrictive	
	Gap	Pronoun	Gap	Pronoun
Definite head:	3	40	32	16
Indefinite head:	4	18	76	84

<div align="center">

Chi-square =   Chi-square =
1.902   5.434
$p < 0.20$ [N.S.] $p < 0.02$

</div>

That is, in nonrestrictive clauses, the definiteness of the head is not statistically significant, whereas it is significant at the 0.02 level for restrictives, with indefinites being more favored in resumptive pronoun clauses than in gap-containing clauses.

Now let us consider gap-containing and resumptive pronoun clauses separately. The figures for gap-containing clauses are shown in 25 and those for resumptive pronoun clauses are shown in 26:

(25) **English:**

	Gap-containing clauses only	
	Nonrestrictive	Restrictive
Definite head:	3	32
Indefinite head:	4	76

<div align="center">

Chi-square = 0.543 $p < 0.50$
[N.S.]

</div>

(26) **English:**

	Resumptive pronoun clauses only	
	Nonrestrictive	Restrictive
Definite head:	40	16
Indefinite head:	18	84

<div align="center">

Chi-square = 45.007 $p < 0.001$

</div>

The pattern that emerges for English in 25 and 26 is strikingly like the one for Yiddish: for gap-containing relative clauses, there is no significant interaction between definiteness and restrictiveness, while, for resumptive pronoun clauses, restrictives tend strongly to be indefinite. That is, there is no significant difference in the likelihood of occurrence of gap-containing clauses like 27a,b, while analogous resumptive pronoun clauses like 28a,b have a very different likelihood of occurrence, indefinites like 28a being far more frequent than definites like 28b:

(27) a. ...I want to show you **some things in the library that you could look at during the vacation time...** (COU:c-786)

    b. There's a big dinner for Reagan a couple of weeks ago and, depending on how much that people paid, they got to have cocktails with him before or an—— and then everybody got to eat with him and then **those who paid a lot more** got to meet with him afterwards... (COU:c-649)

(28) a. My son, God bless him, he married **this girl which I like** *her*. (AK:E.W.) [2]

    b. She sent them in **the working room where my father worked**, you know, *there*. (AK:G.R.)

We shall return below to the exceptional cases, those 16 resumptive pronoun clauses that are restrictive and definite, of which 28b is one.

4.3. The 'file card' account.

The obvious question at this point is: What do nonrestrictives have in common with indefinite restrictives? In what way do they form a natural class? The answer certainly does not lie in any morphosyntactic explanation of definiteness per se, nor does it lie, I believe, in a static truth-conditional semantic account. Rather, I believe that what is needed is a dynamic account of how hearers understand sentences and increment their discourse-models on the basis of those understandings. Intuitively, when a hearer processes a non-expletive NP, s/he must do something with respect to the discourse model. In the typical case, if the NP is indefinite, it represents a 'Brand-new' entity, as in 29a, and the hearer must add that entity, or construct a new file card, following Heim 1983. And, typically, if the NP is definite, it represents something already evoked in the discourse model, as in 29b, or something assumed to be present in the hearer's knowledge-store, as in 29c, in which case the hearer must activate the appropriate existing file card, or else it represents an entity which the hearer is assumed to be able to infer on the basis of prior knowledge s/he is assumed to have, as in 29d, in which case the hearer must construct a file card out of existing material:

(29)   a.   He bought **a house** and a car.
      b.   **The house** is a colonial.
      c.   **Dolly Madison** lived in it for three years.
      d.   But **the roof** is shot.

Now let us reconsider relative clauses. First, as is well known, nonrestrictive relatives do not affect the reference of the head NP. That is, whatever the hearer is doing with respect to the relevant NPs in 29, s/he is doing the same thing with respect to the *heads* alone of the corresponding complex NPs in 30:

(30)   a.   He bought **a house, which, by the way, I had found**, and a car.
      b.   **The house, which is in Society Hill**, is a colonial.
      c.   **Dolly Madison, who he always admired**, lived in it for three
          years.
      d.   But **the roof, which is slate**, is shot.

That is, the hearer activates or constructs the appropriate file cards purely on the basis of the heads, the nonrestrictive clauses representing information which presumably does not yet exist on those file cards and which must be added to them.

      Now consider definite restrictive relative clauses. Typically, the NP that represents the entity whose file card the hearer is being told to activate is the whole complex NP consisting of the head *plus* the relative clause. Unlike the case of nonrestrictives, the relative clause does not simply add some property to an independently selected file card; on the contrary, the relative clause represents information which *must already be* on the file card for the hearer to select that card. Consider 31:

(31)   a.   He bought **a house** in Society Hill and a house down the shore.
      b.   **The house that's in Society Hill** is a colonial.
      c.   **The First Lady who introduced ice cream in America** lived in
          it for three years.
      d.   But **the part of the roof that's over the dormer** is shot.

In 31b, the hearer doesn't add the property of being in Society Hill to the file card for the house; rather, this property is already on one of the two 'house' file cards added for the previous sentence, and it is its presence on that card that induces the hearer to select it in order to add the property of being a colonial. Likewise, the relative clauses in 31c and 31d

are not new attributes but (assumed) old information which the hearer must already have simply to evoke the intended entity.

Thus we see a clear difference in the functioning of nonrestrictive and definite restrictive relative clauses. Interestingly, indefinite restrictive relatives typically function more like nonrestrictives in this respect than like definite restrictives. Consider 32:

(32) a. He bought **a house which he'll move into in June.**
       b. **A realtor that I had recommended** found it for him.

In both cases, the hearer has to add a new file card, as signaled by the indefiniteness of the NPs. And, in each case, the file card to be added need represent only the entity described by the head, the information in the relative clause simply being an additional property of that entity to be noted on the independently constructed file card.

I believe that the difference discussed here between nonrestrictives and indefinite restrictives on the one hand and definite restrictives on the other is the relevant factor for whether a Yiddish or English speaker can felicitously utter a resumptive pronoun relative clause rather than a gap-containing relative clause: gap-containing clauses are always permissible in both languages (island-environments notwithstanding), but resumptive pronouns may occur felicitously just in case the entity evoked by the whole NP is in fact evoked by the head, the relative clause serving simply to predicate some property of that entity, that is, where the appropriate file card has already been independently constructed/activated.

Thus, the file card account predicts that English and Yiddish will have a felicitous resumptive pronoun correlate of, for example, 32a but not of 31b, shown for English in 33:

(33) a. He bought **a house which he'll move into** *it* **in June.** (cf. 32a)
       b. #**The house that** *it***'s in Society Hill is a colonial.** (cf. 31a)

Note that the file card account presented here means that resumptive pronouns work exactly like ordinary discourse pronouns, as argued by Doron 1982 and Sells 1987, inter alia, and it also explains why one finds other anaphoric expressions in the place of resumptive pronouns, e.g. demonstratives, as in 34a, coreferential full NPs, as in 34b, and even referentially related but non-coreferential pronouns, as in 34c, and full NPs, as in 34d:

(34) a. I had **a handout and notes from her talk that** *that* **was lost too.** (AK: Gillian Sankoff)
       b. He's got **this lifelong friend who he takes money from the parish to give to** *this lifelong friend*. (AK: P.)
       c. I have **a manager, Joe Scandolo, who** *we***'ve been together over twenty years.** (AK: Don Rickles, 5/82)
       d. You assigned me to **a paper which I don't know anything about** *the subject*. (AK: Lila Gleitman)

That is, whatever means speakers generally have for referring to already evoked discourse entities will be at their disposal in these resumptive pronoun clauses.

Interestingly, the distinction I am drawing here between definite and indefinite restrictive relatives is similar to one made in McCawley 1982, and it is in fact the same one made in Crain and Steedman 1982, in a very different domain, the processing of garden path sentences. Crain and Steedman show that the expected misparse of garden path sentences often fails to obtain just in case the NP is indefinite. Consider 35:

(35) a. **The teachers taught by the Berlitz method** passed the test.
      b. **The children taught by the Berlitz method** passed the test.
      c. **Teachers taught by the Berlitz method** passed the test.
      d. **Children taught by the Berlitz method** passed the test.
          (= Crain and Steedman 1982, ex. 16)

In spite of the different plausibility levels, 35c,d were misparsed less often than 35a,b (p < 0.001). While garden path phenomena are not directly relevant to resumptive pronoun clauses, Crain and Steedman's experimental evidence strongly supports the claim that hearers process indefinite and definite restrictive relatives differently. (In fact, their explanation for the garden path results is directly compatible with what has been argued above for the resumptive pronoun data.)

4.4. The 'concept' claim revisited.
Although Sells' (1987) 'concept' claim, that resumptive pronouns may not have nonspecific reference, is not made explicitly for English, he does suggest (pp. 288f.) that it holds also for English. While I cannot speak for the situation in Hebrew, I should like to return now to why someone might think that English—or Yiddish—lacked a 'concept' understanding for resumptive pronouns.
  Briefly, the examples given in the literature of infelicitous uses of resumptive pronouns with 'concept' understandings involve definite NPs with restrictive relative clauses, e.g. 36:

(36) a. Every man will eventually find **the woman that he seeks**. (=Sells 1987, ex. 49a)
      b. Every man will eventually find **the woman such that he seeks *her***. (= Sells 1987, ex. 49b)

The claim, attributed to Steven Weisler, is that 36b cannot have the 'concept' understanding. While Sells notes 'a potential problem' with the definite article in 36b, he explicitly ignores it, suggesting that only definites can have 'concept' interpretations. Of course, if 'concept' reference is akin to 'nonspecific', 'nonreferential', etc., then it may of course be indefinite. In any event, as noted above, both the Yiddish and the English corpus contain nonspecific heads with resumptive pronouns, as shown in 14a and 15 above. The difference between them and 36b is that, in 14a and 15, the relative clause represents a property of some independently evocable entity, which happens to be nonspecific, whereas, in 36b, the relative clause is a defining property of that entity. Let us look more closely at 14a, repeated here for convenience as 37:

(37) [A messenger comes to tell a Jewish community that a priest wishes to have a Bible contest with the Jews.]
      di shtetl zol shteln **a yidn vos *er* zol geyn oysfregn dem galekh,**
      the village shall place **a Jew that *he* shall interrogate the priest,**
      un der galekh zol oysfregn dem yidn. (RP:220)
      and the priest shall interrogate the Jew.
      'The village should pick **a Jew who *[he]* will question the priest...**'

In file card terms, the hearer of 37, upon hearing the head *a yidn* 'a guy', adds a new file card for this new, arbitrary guy. Then, upon hearing the relative clause, the hearer adds to that file card the information that this arbitrary guy will interrogate the priest. Note the difference in the function of the relative clause in 37 and in 38:

(38) a. men zukht **dem yidn vos** *[e]* **zol geyn oysfregn dem galekh.**.
one seeks the guy that *[e]* **shall go interrogate the priest.**
b. #men zukht **dem yidn vos** *er* **zol geyn oysfregn dem galekh.**
one seeks the guy that *he* **shall go interrogate the priest.**

In 38, where the file card is selected because it already has the information on it that this arbitrary guy will go interrogate the priest, the resumptive pronoun is not felicitous. Thus, while I do not know what the actual situation is in Hebrew, it is clear that in Yiddish and English there is nothing incompatible between 'concept' interpretations and resumptive pronouns, although the particular examples of 'concept' interpretations that are typically invented are infelicitous with resumptive pronouns for independent reasons.

4.5. The exceptions.
At this point, the question remains why there are so many exceptions in the English corpus to the predictions made by the file card account, that is, why there are 16 tokens of resumptive pronoun clauses which are both definite and restrictive, as shown in the figures in 26, one of which is given in 28b.

I believe one answer is that the English corpus is a corpus of spontaneously produced data and as such is likely to contain dysfluencies and other performance errors. In fact, the resumptive *there* in 28b clearly has the ring of an afterthought. Note in this vein that resumptive pronouns have already been established as arising from poor syntactic planning, in cases where they occur in what would be inaccessible extraction sites. Of course, there is no reason to believe that the only time a speaker cannot go through with an intended extraction is when it is ruled out or strongly disfavored by the syntax; very plausibly, speakers may fail to perform an intended extraction for extralinguistic reasons—memory lapse, distraction—whatever reasons lead to dysfluency in simple sentences. In fact, 4 of the 16 are datives, as mentioned above, an argument that is universally difficult to extract.

Interestingly however, a close look at the remaining cases in question suggests that not all are the result of dysfluency: 6/16 are cases where the head is introduced by some form of *that*, e.g. 39:

(39) a. In the Preservation News they had a feature article on the replastering of **those incredible figurines that** *they* **look like they're holding up the balcony.** (AK:C.R.)
b. I know it can't be X. Do you know **that guy who Mommy's typing for** *him*? (AK: D.K.)
c. He's very good at **those gold leaf letters you put** *them* **on from the inside.** (AK: T.M.)

Although these are all technically definite, in fact they act like indefinites in at least one other way: they are amenable to extraposition, usually reserved for indefinites (Ziv and Cole 1974, Ziv 1975), as in 40:

(40) a. I saw **those incredible figurines** yesterday **that look like they're holding up the balcony.**
b. **That guy** just called **who Mommy's typing for.**
c. I found **those gold leaf letters** in the bookstore **that you put on from the inside.**
d. **Those people** may come **whom you want.** (= Guéron 1980, ex. 111b)

Furthermore, and perhaps relatedly, another 6/16 of the definite restrictives with resumptive pronouns in the English corpus are of the form Demonstrative-BE-NP and 1/16 is of the form Pronoun-BE-NP, e.g. 41:

(41) a. Was that **the girl that some wines made *her* feel sick?** (AK:P/S)
    b. Was that **the one that—** *she* **teaches at Temple?** (AK: H.D.)
    c. This is **the guy that** *he* **stabbed Shields and gave him 64 stitches across his chest**, right? (AK:M.)

Interestingly, in all these cases, the hearer is assumed to already have the appropriate file card activated: in 41a,b, the speaker is asking the hearer to check the currently activated file card to see whether it has certain information, so that the speaker can determine whether s/he in fact has the right file card activated. Similarly, 41c presumes that that hearer has already activated the appropriate file card and is asking for confirmation, presumably for rhetorical effect. Thus, in all these cases, the hearer is not expected to activate the file card on the basis of the information in the relative clause. Notice that the complex NPs in 41 can be paraphrased by free relatives, rather rare in Modern English when the complementizer is *who*; these are shown in 42:

(42) a. Was that **who some wines made feel sick?**
    b. Was that **who teaches at Temple?**
    c. This is **who stabbed Shields...,** right?

In addition, two of the sixteen definite restrictives with resumptive pronouns have *same* in their head; one of them, shown in 43a, clearly works like an indefinite; the other, shown in 7b, is a fragment and its behavior cannot be determined. In addition, one of the five tokens of the form Demonstrative-BE-NP, given in 43b, likewise does not work like a definite in the usual way:

(43) a. The black Mafia, I believe, is run by **the same Mafia that everybody else knows** *it* **as the Italian Mafia or the Cosa Nostra...** (AK:R.R.)
    b. That's **the only dream that** *it* **came true.** (AK:D.)

In 43a, we are told that the black Mafia is run by a(nother) Mafia and that this other Mafia has the property that it is known as the Italian Mafia. Notice that such definite NPs with *same* can occur freely in *there*-sentences, environments for (conceptually) indefinite NPs, as in 44:

(44) a. Last year I went to BLS and CLS. There were **the same people at both conferences.**
    b. There are **the same people at this BLS that I saw at the last BLS.**

Note that, in 44b, we even have a definite NP with a restrictive relative in a *there*-sentence: what the relative is doing is not telling the hearer which file card to activate but rather giving information about the filecard activated by the head *people at this BLS*, to wit, that they were seen at the last BLS.

    Analogously, in 43b, the definite head does not work in the stereotypic definite way: all that the complex NP is doing is conveying the information that no dream other than the one already under discussion has the property of having come true and, by conventional implicature, that the one under discussion has.

Finally, we are left with two unexplained definite restrictive resumptive pronoun clauses, the dative sentence fragment in 7a, where we have insufficient information, and the full sentence in 45:

(45)   Let's go to the library and get the Babar book, get **the book that Beto's going to read** *it*. (AK:J.B.)

This turns out to be one of the most compelling pieces of evidence for the importance of context in discourse studies. If we just considered the clause in which the relative clause occurs, we would be at a loss to explain it. With the prior context, however, the mystery is solved: the speaker, a young woman speaking to the children for whom she is babysitting, has already evoked the appropriate filecard with the NP *the Babar book*; the repetition containing the relative clause serves perhaps simply a rhetorical purpose in that it articulates something the hearers presumably already know: that this book has the property that Beto is going to read it. What it is *not* doing is telling them which book is under discussion.

Thus we see that the exceptions are not really exceptions; rather the morphological marking of definite/indefinite has only an imperfect correlation with the cognitive processes with which we associate them, and it is these cognitive processes, not their morphological markings, that are reflected in the use of resumptive pronoun clauses in Yiddish and English.[3]

Notes

[*]Short though it may be, this paper owes a great deal to many people. Among others, I heartily thank A.S. Kroch for the use of his resumptive pronoun corpus, M. Benya, N. and A. Meilijson, and M. Schaechter for their Yiddish intuitions, S. Cote, N. Erteschik-Shir, K. Lambrecht, R. Oehrle, M. Steedman, S. Steele, E. Vallduví, L. Walker, and B.L. Webber for their invaluable advice and assistance, and the Department of Computer and Information Science, University of Pennsylvania, for excellent computational resources.

[1]This transcript also contains three resumptive pronoun clauses, all of the island type and therefore not used in this study.

[2]I am counting such instances of *this* as indefinite; cf. Prince 1981.

[3]Other syntactic forms with closely related, though not identical, discourse functions, are missing complementizers with extracted subjects, as in i, apparent 'run-ons', as in ii, extraposed relative clauses, as in iii, and *kind*-clauses, as in iv. See Prince, In prep.

i.   I have a friend of mine in the history department **teaches two courses per semester.** (= Lambrecht 1988, ex. 9)
ii. There are many Americans **they approve of violence.** (= Lambrecht 1988, ex. 37c)
iii. I read a book during the vacation **that was written by Chomsky.** (= Guéron and May 1984, ex. 17a)
iv. He was one of those comedians **that he could make the whole audience laugh without even saying a word.** (Barry Reisman, WIBF-Philadelphia)

References

Ariel, M.1990. *Accessing NP antecedents*. London: Routledge.

Bever, T., Carroll, J., and Hurtig, R. 1976. Analogy or ungrammatical sequences that are utterable and comprehensible are the origins of new grammars in language acquisition and linguistic evolution. In Bever, T., Katz, J., and Langendoen, D.T., eds. *An integrated theory of linguistic ability*. NY: Crowell. Pp. 149-82.

Crain, S. and Steedman, M. 1982. On not being led up the garden path: the use of context by the psychological syntax processor. In Dowty, D., Karttunen, L., and Zwicky, A., eds. *Natural language processing: psychological, computational, and theoretical perspectives*. ACL Studies in Natural Language Processing. Cambridge: Cambridge Univ. Press. Pp. 320-58.

Doron, E. 1982. On the syntax and semantics of resumptive pronouns. *Texas Linguistic Forum* 19.1-48.

Engdahl, E. 1979. The Nested Dependency Constraint as a parsing principle. In Engdahl, E. and Stein, M.J., eds. *Papers presented to Emmon Bach by his students*. Amherst, MA: Univ. of Massachusetts.

Erteschik-Shir, N. 1989. Resumptive pronouns in islands. Presented at the Conference on the Psycholinguistics of Island Constraints. Univ. of Ottawa. September 16, 1989.

Guéron, J. 1980. On the syntax and semantics of PP-Extraposition. *Linguistic Inquiry* 11.637-78.

Guéron, J. and May, R. 1984. Extraposition and logical form. *Linguistic Inquiry* 15.1-31.

Heim, I. 1983. File change semantics and the theory of definiteness. In Bauerle, R., Schwarze, C., and von Stechow, A., eds. *Meaning, use, and the interpretation of language*. Berlin: Walter de Gruyter. Pp. 164-89.

Kayne, R. 1981. ECP extensions. *Linguistic Inquiry* 12.93-134.

Keenan, E.L. and Comrie, B. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8.63-99.

Kroch, A.S. 1981. On the role of resumptive pronouns in amnestying island constraint violations. In *CLS17*. Chicago: Univ. of Chicago. Pp. 125-35.

Lambrecht, K. 1988. There was a farmer had a dog: syntactic amalgams revisited. In *BLS 14. Papers from the Parasession on Grammaticalization*. Berkeley, CA: Univ. of California.

Langendoen, D.T. 1970. The accessibility of deep structures. In Jacobs, R. and Rosenbaum, P., eds. *Readings in English transformational grammar*. Boston: Ginn. Pp. 99-106.

Maling, J. and Zaenen, A. 1980. A base-generated account of Scandinavian extraction phenomena. In Pullum, G. and Jacobson, P., eds. *On the nature of syntactic representation*. Dordrecht: D. Reidel.

McCawley, J.D. 1982. The syntax and semantics of English relative clauses. *Lingua* 53.99-149.

Olsvanger, I. 1947. *Royte pomerantsn*. NY: Schocken Press.

Sells, P. 1987. Binding resumptive pronouns. *Linguistics and Philosophy* 10.261-98.

Prince, E.F. 1981. On the inferencing of indefinite-*this* NPs. In Joshi, A.K., Webber, B.L., and Sag, I.A., eds. Elements of discourse understanding. Cambridge: Cambridge Univ. Press. Pp. 231-50.

Prince, E.F. In prep. Resumptive pronoun clauses and related phenomena in English and Yiddish discourse.

Ziv, Y. and Cole, P. 1974. Relative extraposition and the scope of definite descriptions in Hebrew and English. *CLS 10*. Pp. 772-86.

Ziv, Y. 1975. On the relevance of content to the form-function correlation: an examination of extraposed relative clauses. *CLS 11. Parasession on Functionalism*. Pp. 568-79.

On Predicting Pragmatic Relations

John F. Richardson   Alan W. Richardson
Baruch College   Northwestern University

1. Notes on the Modern Muddle in Pragmatic Theory

There is a peculiar tension in the pragmatics literature. Over the last several years, empirical work has demonstrated the true lexical, syntactic, semantic and discourse-theoretic complexities of individual pragmatic phenomena. For instance, Horn (1978) noted that there are lexical exceptions to the otherwise (largely) pragmatically explicable phenomenon of neg-raising, as shown in (1).

1.a. I don't want to die. (I want not to die.)
   b.*I don't hope to die. (I hope not to die.)

Horn (1989) has also shown that precisely those *wh*-questions that are used rhetorically as universal negatives allow certain bits of syntax otherwise marginal to impossible in interrogative clauses, but allowed in universal negative constructions, as shown in (2).

2.a. Who (but a total idiot) would say that?
   b. I found out who (*but a total idiot) said that.
   c. No one (but a total idiot) would say that.

Horn (1984) has also argued for the existence of pragmatically motivatable but nonetheless truly semantic privative ambiguities, such as the one that explains the noncontradictory nature of (3).

3. There are three cows in the field; a bull and two cows.

Finally, Morgan has (1978) postulated "short-circuited" implicatures or conventions of use in order to preserve a pragmatic account of why questions like (4a) can convey a request, while acknowledging that synonymous questions like (4b) can do so only much less directly.

4.a. Can you hand me the paper?
   b. Are you able to hand me the paper?

At the same time that the collective weight of these examples and many others like them should have established that pragmatic theory must be internally complex and that even a realistically complex pragmatic theory will have to be able to interact with the other components of linguistic theory in various complicated ways, a (meta)theoretical movement informally known as radical pragmatics has been remarkably successful in convincing many that it is desirable to reduce all of pragmatic theory to a few grand principles and that even such a highly simplified pragmatic theory can radically extend its domain of explanation at the expense of the other components. Although it is impossible to pin down the guiding tenets of radical pragmatics unequivocally (as they are seldom actually propounded as such and many of them are not unanimously upheld even within the radical pragmatics community), we feel that the propositions presented in (5) capture its basic attitudinal drift.

5.   A. Rational Reductionism: (1) Pragmatics is the general theory of human rationality applied to language use. (2) Rationality stands in stark opposition to convention and arbitrariness. (3) Rationality can explain particularized and generalized implicatures equally well. (4) Rationality can serve as the basis of a truly predictive and not merely restrictive pragmatic theory.

B. Semantic Minimalism: (1) Pragmatic theory allows one to posit austerely truth conditional logical forms. (2) Pragmatic theory minimizes the number of actual ambiguities that must be posited. Indeed, any putative ambiguity that can be given a pragmatic explanation is thereby shown to be spurious. (3) The only semantic properties relevant to pragmatic analysis are the entailment relations that hold between uttered and strategically unuttered propositions.

C. Lexicogrammatical Minimalism: (1) Only a few, utterly straightforward aspects of surface syntax like consituent order and (relative) complexity are ever relevant to pragmatic analysis. (2) These few aspects of grammar are themselves relevant only if two pragmatically and syntactically dissimilar sentences are synonymous. Otherwise, the pragmatic difference should be pinned on the semantic difference. (3) The only exception to these "rules" is conventional implicature, which is a notion to be invoked sparingly and only as a property of lexical items, not of entire grammatical constructions.

D. Pragmatic Minimalism: One can account for all pragmatic phenomena using fewer maxims than those proposed in Grice (1989), perhaps even enfolding the co-operative principle into one of the few remaining maxims.

We will define our own views against these propositions. Briefly, we will argue that important details of most if not all generalized implicatures fall well beyond the meager predictive powers of radical pragmatics, that literally none of the propositions in (5) deserve unequivocal support and that most deserve outright rejection. We will begin with a critique of rational reductionism and then critique the "three minimalisms" in turn.

2. Rational Reductionism

Let us first consider the proposition that pragmatics is the general theory of human rationality applied to language use. Although this part of the rationality thesis looks like it promises the ultimate in deep explanation, no pragmatician has ever proposed or endorsed a general theory of human rationality that is remotely well articulated enough to give this claim any real content. This, in turn, immediately robs the next three propositions of any real content, since one can hardly pretend to know what rationality stands in opposition to, how truly predictive a concept it is or what range of phenomena it can be used to explain until one has some semblance of an account of what rationality actually is. But if we charitably assume that we can intuit rationality well enough to evaluate the plausibility of these claims, things actually get worse. For it is intuitively true that there is often more than one perfectly rational means for achieving a given goal G in a given context C. Moreover, when such a situation arises, it is always rationally permissable and frequently rationally advisable to arbitrarily pick one of the available means as the conventional way to achieve G-like goals in C-like contexts. To test this, simply consider the question of which side of the road one should drive on. There is certainly no rational basis for preferring the British to the American system or vice versa. But it is highly rationally advisable to simply choose one or the other system and to conventionalize whatever choice is made.

This finishes all hope of putting rationality into stark opposition to either convention or arbitrariness. But this is old news in the linguistic world; consider the phenomena in (1) - (4), Sadock's (1978; 1984b) work on calculable idioms and rational but unpredictable compounds or Lakoff's (1987) concept of motivated extensions to grammatical constructions. All we have added to the discussion is some intuitive

grounding in naive rationality theory that should help remove the apparently widespread feeling that there is something less than satisfying or even paradoxical about the notion of rationally motivated arbitrary conventions.

Of course, the best way to force a nonarbitrary choice between two otherwise equally rational practices is to define either the goal or the context in question in such detail that ultimately only one of the practices remains fully rationally defensible. This means that if pragmatics is simply rationality theory applied to language use, implicature should be highly dependent on the minute details of communicative goals and contexts. But while this is in fact true by definition for particularized implicature, it is false by definition for generalized implicature. This suggests that the third proposition of the rationality thesis is particularly misguided. Indeed, we submit that generalized implicature is far too insensitive to the details of actual communicative goals and contexts to be explained in any depth by rationality theory.

Finally, the last proposition under the rationality thesis promises a pragmatic theory that is predictive and not merely restrictive. If, however, pragmatic theory is based strictly on rationality, it can only be a restrictive theory. That is, while certain triplets of context, utterance and intended meaning can surely be ruled out as irrational, there will often be too many rationally defensible possibilities left over to make any determinate predictions. Indeed, it is empirically obvious that particularized implicatures can at best be rationally restricted and not fully predicted (by analysts or interlocutors). But it is precisely because of its merely restrictive nature that rationality is incapable of explaining anything so specific and automatic as a generalized implicature. However, it will become clear below that at least some of the specific and automatic details of at least some generalized implicatures require partially stipulative treatments. Since any appeal to stipulation is by definition a failure of prediction, there would appear to be no subdomain of pragmatic inquiry that is fully within the grasp of a truly predictive theory.

3. Semantic Minimalism

It may be tempting to believe that propositions (1) and (2) of semantic minimalism form the entire rationale for doing pragmatics in the first place. Note, for instance, that Grice's original program was motivated in large part by his desire to extend the traditional analyses of the logical connectives, quantifiers and so forth to their natural language analogues, with pragmatics supposedly bridging the gap between the communicatively rich intuitive meanings of the latter and the austerely truth-conditional formal definitions of the former. He also proposed a principle, his modified Occam's razor, which enjoined one from postulating senses beyond necessity and this, too, seemed quite fundamental to his thinking.

Several things should be kept in mind, however. First, Grice correctly realized that parsimony considerations can only serve as a useful guide when one has some cogent basis for deciding just what is, in fact, really necessary - an open ended issue as analytical possibilities come and go. Modern attacks on supposed excesses in sense-postulation seldom even approach the sophistication of Grice's cautiously announced proposal. Turning to Grice's logical traditionalism, note first that Grice was writing as a great innovator. The typically overly optimistic goals and pronouncements of true innovators follow almost automatically from the naiveté that they unavoidably bring to their newly established fields. Those of us who later enter maturing fields, on the other hand, are only as naive as we let ourselves be. Second, Grice allowed himself considerable pragmatic machinery with which to try to bridge the gap between intuitive and posited meanings, whereas radical pragmaticians have been eagerly scrapping much of Grice's machinery while frequently allowing gaps between intuitive and posited meanings greater than anything Grice proposed - a geometrically more ambitious program. Third, the logical analyses Grice wished to uphold had proven their value to semantic analysis in several centuries worth of logical work. The eminently reasonable idea that semantic analyses

should be motivated above all by their value to semantic theory appears, incredibly, to have few adherents among modern pragmaticians.

Consider Horn's (1972) famous claim that the numeric quantifiers strictly mean "at least n" and only implicate "exactly n" via a standard quantity implicature. Sadock (1984a) has pointed out that Horn's semantic proposal can make no sense of arithmetic statements like (6a). Horn (1989) has acknowledged the problem, but, remarkably unfazed by its devastating force, has continued to maintain the earlier analysis. Lest one hope that an appeal can be made to something concerning the technical uses of language or some subtle difference between the number names versus numerical quantifiers, we present (6b) as a perfectly colloquial and clearly quantificational sentence beyond the semantic reach of Horn's proposal.

6.a. The square root of nine is three.
    b. I took six cigarettes with me, gave one to Fred and two to Ed, so I still have three.

A less minimalistic analysis that posits an ambiguity between the "at least n" and "exactly n" readings would not be troubled by (6). Horn (1989) dismissed a version of such a proposal by Kempson (1986) and allies by pointing out that this would create an infinitude of ambiguous terms. This attempt at a crushing parsimony argument is badly misconstructed. One does not, for instance, refute the phrase structure ambiguity shown in (7) by pointing out that it would render infinitely many NPs structurally ambiguous.

7. [[NP and NP] and NP]   versus   [NP and [NP and NP]]

Closer to our present numerical home, the fact that every ordinal number term starting with *third* also has a portional reading, as seen in (8), does not cry out for a pragmatic account.

8. The six hundred and twenty-third patron receives one six hundred and twenty-third as much as the first.

In short, Horn has confused the unimportant question of how many items are "in the domain" of an ambiguity with the important question of how much apparatus is required to handle the ambiguity. There is no reason to believe that an "at least n" - "exactly n" ambiguity would require much machinery at all.

Now consider another classical quantity implicature. Perhaps no pragmatic analysis is more popular, or less radical feeling, than the one that derives the exclusive reading of *or* from the inclusive reading by the quantity scheme in (9).

9.a.[P and Q] unilaterally entails [P or-incl Q].
   b.Hence, if one can truthfully utter [P and Q], one should not utter [P or-incl Q].
   c.Hence, if one does utter [P or-incl Q], one thereby implicates [not [P and Q]].
   d.The conjunction of what is said, [P or-incl Q], and what is implicated, [not [P and Q]], is truth conditionally equivalent to [P or-excl Q].

Unfortunately, this apparently explanatory analysis mispredicts in the case of nonbinary coordination. The case of ternary coordination would, for instance, have to follow the scheme in (10).

10.a. [P, Q and R] unilaterally entails [P, Q or-incl R].
   b. Uttering the latter should implicate the negation of the former.
   c. What is conveyed should then be what is said plus what is implicated,
     i.e. [[P, Q or-incl R] and [not [P, Q and R]]].

But this predicts that (11) should convey that diners can choose any one or any two of the listed desserts, just not all three. In fact, however, it conveys that diners may choose one and only one dessert.

    11. Diners may have cake, pie or ice cream.

A slimmed-down version of our sincerest effort to preserve the spirit of the traditional quantity implicature is given in (12).

    12.a. Assume prototypical coordination to be binary.
       b. In prototypical case, generate implicature as metalinguistic quantity
         implicature as follows:
         a'. [P and Q] is true if and only both conjuncts are true.
         b'. [P or-incl Q] is true if and only if at least one conjunct is true.
         c'. For essentially the same reasons as those given in (9), [P or-incl Q]
           will tend to implicate that not both of its conjuncts are true.
       c. The implicature in (c') is, in the prototypical case, equivalent to saying
         [P or-incl Q] implicates that exactly one of its conjuncts is true. Assume
         that the implicature is reanalyzed in just this way.
       d. Extend the reanalyzed version of the implicature to all nonprototypical cases.

If this is correct, several very unminimalistic devices must be adopted. First and most obviously, prototype structures must be posited even for so "logical" a portion of the vocabulary as the conjunctions. Second, these prototype structures must have an abstract, schematic reality separate from the instantiations of the conjunctions in individual sentences, since it would be silly to claim that, for instance, (11) is prototypically a case of binary coordination, as it is obviously a case of ternary coordination plain and simple. Hence, third, the implicature does not directly exploit the entailment relations between uttered and unuttered propositions, but the relative lexical semantic "strength" of the two conjunctions. Fourth, we must posit an implicature reanalysis rule and a process by which the output of this rule is extended to nonprotypical cases. Finally, we at least strongly suspect that (12) really makes more sense as the pragmatical motivation of a privative ambiguity than as a "pure" conversational implicature as classically conceived.

4. Lexicogrammatical Minimalism

The relevance of constituent order to implicature has been argued on the basis of the intuitive nonsynoymy of pairs of sentences like those in (13).

    13.a. Mary went to New York and bought some records.
       b. Mary bought some records and went to New York.

The relevance of relative lexicogrammatical complexity to pragmatics can be seen in the intuitive nonsynonymy of pairs of sentences like those in (14).

    14.a. Mary likes to go to New York and buy records.
       b. Mary likes to go to New York and to buy records.

Since the order of the two VPs is the same in (14b) as it is in (14a), the failure of (14b) to implicate that the two VPs denote temporally consecutive parts of a complex action can only be traced to the occurence of the second infinitive marker. A reasonable account would be that in using two full infinitive phrases instead of one, a speaker would be going out out of her way to make the actions denoted by the two VPs appear separate and unrelated, thereby preempting the implicature in (14a). Evidence for the pragmaticity

of the phenomenon can be seen by considering (15a). Since there is no option of using just one gerund marker, as in (15b), uttering (15a) would not count as going out one's way to make the actions appear independent. Hence the implicature is fully available (though not, of course, mandatory).

15.a. Mary enjoys going to New York and buying records.
   b.*Mary enjoys going to New York and buy records.

Note now that the implicature is available when units larger than VPs such as Ss are conjoined, as seen in (16).

16.a. Mary went to New York and she bought some records.
   b. Mary bought some records and she went to New York.

Now, it is not immediately clear why the addition of a truth conditionally irrelevant subject pronoun in the second clause does not preempt the implicature just as the extra infinitive marker did in (14b). Note, however, that conjoining two that-clauses instead of two simple sentences can sometimes weaken the implicature, as in (17B), though this, too, is a palpably weaker effect than that in (14b) and one needs to use some real excess verbiage as in (17B') to get a comparably robust phenomenon.

17. A: Is it true that Mary joined Wicca and Ted left her?
    B: It is true that Mary joined Wicca and that Ted left her.
    B': It is true that Mary joined Wicca and it is true that Ted left her.

The question at this point is how much is really being explained by the claim that relative lexicogrammatical complexity has pragmatic effects versus how much we will end up having to stipulate concerning what particular bits of lexicogrammatical complexity have what particular pragmatic effects (if any).

But things get more complicated still, since the pragmatic effects in turn have grammatical effects. Consider the apparent coordinate structure constraint violations in (18), based on examples discussed in Lakoff (1986).

18.a. It's the kind of store you can go to and buy almost any record.
   b. It's the kind of record you can go to almost any store and buy.

Lakoff treated first conjunct violations such as (18a) as fundamentally distinct from second conjunct violations such as (18b). We are, however, unaware of any motivation for such a dichotomy and would propose, as a first stab, that coordinate VPs whose denoted actions are implicated to be ordered parts of a single complex action allow unilateral extraction from either conjunct.

This contrasts with the case in (19), also discussed by Lakoff as well as by Goldsmith (1985), where extraction is allowed only in the first conjunct.

19.a. How many courses can you teach and (still) publish at least two papers?
   b.*How many papers can you teach four courses and (still) publish?

The crucial difference in intuitive meaning governing the grammatical difference between the sentences in (18) and (19) can be seen in (20) - (23). In sentence (20a), the action denoted by the second of the conjoined VPs flows naturally from the first. This gives rise to the basic ordered complex action implicature, with the grammatical consequence that extraction is allowed from either conjunct, as seen in (21).

20.a. Bart took a year off and worked on his book.
   b. Only an iron man could stay up all night and run a marathon the next day.
   c. The Japanese eat tofu and don't get cancer.

21.a. How long did Bart take off and work on his book?
   b. Which book did Bart take a year off and work on?

In sentence (20b), the action denoted by the second conjunct is also taken to follow that denoted by the first, but the whole point of the sentence is that the second action does not flow naturally from the first, but is unexpected. In this case, only extraction from the first clause is permissable, as seen in (22).

22.a. How late can you stay up and (still) run a marathon the next day?
   b.*How many miles can you stay up all night and (still) run the next day?

Sentence (20c) can be read either way, depending on whether eating tofu is taken to help or hurt one's chances against cancer. But note that while (23a) is "ambiguous" like (20c), (23b) has only the "natural flow" reading, just as expected.

23.a. What do the Japanese eat and not get cancer?
   b. Which disease do the Japanese eat tofu and not get?

Of course, when two conjoined VPs denote actions or states have nothing to do with each other, the coordinate structure constraint holds for both conjuncts, as seen in (24).

24.a. Violetta raises horseshoe crabs and plans to work in television.
   b.*What does Violetta raise and plan to work in television?
   c.*What does Violetta raise horseshoe crabs and plan to work in?

So we appear to have yet another case of pragmatically governed syntax. The question that now arises is whether we don't in fact have to simply acknowledge the existence of three separate coordinate VP constructions, each with its separate syntax and, crucially, separate conventionalized pragmatics. For note now that the coordinate structure constraint holds exceptionlessly in sentence coordinations, even when the sentences have the same semantics and pragmatics as the exceptional VP coordinations just discussed. Consider (24) and (25).

24.a. They went to the store and (they) bought some records.
   b. Which store did they go to and (*they) buy those records.
   c. Which records did they go to the store and (*they) buy?

25.a. They taught five courses and (they) published three papers.
   b. How many courses can they teach and (*they) publish three papers?

The interplay between syntax and pragmatics is thus extremely complex in this case. Surely more than just constituent order and mere relative lexicogrammatical complexity must be referred to in accounting for all the goings-on here. In particular, it seems that conventionalized pragmatics does indeed extend to units of grammar of a significantly higher order than mere lexical items. But while lexical conventional implicature is a purely stipulative matter, the conventionalized implicatures associated with entire constructions are of a sort whose existence we have been arguing for throughout: fairly easily rationalizable inferences broadly consistent with but by no means fully determined by general principles.

4. Pragmatic Minimalism

While all the foregoing is highly important, ground zero of radical pragmatics is no doubt the modern search for a mere handful of maxims capable of all the work that Grice's four maxims and nine submaxims purported to do. The most extreme version of pragmatic minimalism is to be found in the "relevance" framework of Sperber & Wilson (1986), which supposedly uses only a single "principle" of relevance. We have, however, been unable to determine what, if anything, the content of Sperber & Wilson's claim is and will concentrate instead on the far more intelligible claims by Horn (1984; 1989) and fellow travellers.

Horn proposed to reduce all of pragmatic theory to three grand maxims. The first was apparently equivalent to Grice's quality maxim, though Horn did not articulate it. Although we believe that some important issues hinge on how Horn proposes to state this maxim, we will ignore these issues as for now unbroachable. Horn spared no effort, however, in articulating and defending his other two maxims, which he named Q and R in honor of quantity and relevance, but which in fact supposedly "collected" all of Grice's non-quality maxims. The principles are shown in (26).

26. The Q-Principle: Make your contribution sufficient. Say as much as you can.
    The R-Principle: Make your contribution necessary. Say no more than you must.

As noted in Levinson (1987), Horn is equivocally quantifying over informational content and mere verbiage in these principles. For instance, Horn claims that the R principle covers both Grice's second quantity maxim and third manner maxim, as reproduced in (27).

27. Quantity 2: Do not make your contribution more informative than is required.
    Manner 3: Be brief.

Levinson appears somewhat ambivalent about what to think about this. But note that the famous inference from (28a) to (28b), which Horn taxonomizes as an R-based implicature, is a case in which the utterance of a more verbose but less informative sentence implicates what could be asserted by uttering a less verbose but more informative sentence. This suggests a clash between Grice's second quantity and third manner maxims and refutes Horn's claim that they can both be covered by one grand principle.

28.a. Mary was able to solve the puzzle.
    b. Mary solved the puzzle.

Of course, as we already stated under lexicogrammatical minimalism in (5), pure conversational implicature can only be sensitive to such purely formal matters as relative verbosity precisely when the added verbosity does not add to the information content of a sentence, which fact renders Horn's attempt to equate such necessarily non-correlating formal and informational notions of linguistic quantity highly enigmatic.

This form-content equivocation is not the only one in Horn's work. In his (1984) paper, he claimed that his Q and R principles were somehow involved in the tension between respectful politeness (or aloofness) versus friendliness (or imposition), as seen in (29). Although there certainly is a dynamic tension between these two kinds of politeness, it is hard to imagine what any of this has to do with either the formal or the informational versions of his Q and R principles.

29. Excuse me, er, Professor ... , er, Lar- ..., er, Mr. Horn.

Now, the classical quantity implicatures all involve relative information content, not verbosity or aloofness or anything else Horn may have wished to account for with his theory. Within the restricted domain of relative information content, Horn's Q and R principles are strongly akin to Atlas & Levinson's (1981) principles of quantity and informativeness. Like Atlas & Levinson's dichotomy, Horn's is based on yet another bit of confusion, this time between the Gricean notions of what is implicated versus what is conveyed.

First consider the supposedly opposing pragmatic inference schemes in (30), drawn more or less directly from Horn's work, and their supposed instantiations in (31), which are the derivation of the exclusive reading of *or* and conditional perfection, respectively.

30.   Q-Based Implicata   R-Based Implicata
              S entails W   S entails W
         "W" implicates not S   "W" implicates S

31.a. [P and Q] entails [P or Q];
        "[P or Q]" Q-implicates [not[P and Q]]
     b. [Q iff R] entails [if R, Q];
        "[if R, Q]" R-implicates [Q iff R]

But now consider the alternative inference schemes in (12).

32.a. [P or-excl Q] entails [P or-incl Q];
        "[P or-incl Q]" R-implicates [P or-excl Q]
     b. [not[if [not R], [not Q]]] entails [if R, Q];
        "[if R, Q]" Q-implicates [if [not R], [not Q]]

What this shows is that whether an inference instantiates the Q or R scheme depends strictly on what one takes to be the relevant entailing proposition. Since exclusive disjunction itself entails inclusive disjunction, why not just make this implicature an R implicature? Similarly, since a conditional is (classically) entailed by the negation of the converse of its contrapositive (or whatever that thing is in (32b)), why not derive the biconditional sense from the coordination of what is said and what is Q-implicated? If, however, one sets [P] in the above schemes equal to [not R], all of the equivalences in (33) hold according to classical logic. While some of these equivalences can and have been questioned, they do provide grounds to suspect that the implicatures giving rise to the exclusive sense of or and to conditional perfection are in anything but stark opposition, contrary to Horn's and Atlas and Levinson's repeated assumption.

33.a. The said [[not R] or-incl Q] equals the said [if R, Q]
     b. The unsaid strong prop [[not R] and Q] equals the unsaid
        strong prop [not[if [not R], [not Q]]]
     c. The implicated neg of strong prop [not[[not R] and Q]] equals
        the implicated neg of strong prop [if [not R], [not Q]]
     d. The conveyed [[not R] or-excl Q] equals the conveyed
        [Q iff R] (each equals coordination of said with implicated)

Indeed, the equivalences in (33) were overtly exploited in Prince's (1982) analysis of conditional perfection, which detoured through the exclusive or quantity implicature! Note that a similar confusion can be seen in Horn's attempt to treat examples like (34) as Q implicatures and examples like (35) as R implicatures. While there is without question a complementarity to these examples, the fact is that (34b) entails (34a) just as surely as (35b) entails (35a), so no entailment-based dichotomy will shed any light its nature.

34.a. I found a finger.
   b. I found someone else's finger.

35.a. I broke a finger.
   b. I broke one of my own fingers.

     What Horn, Atlas and Levinson have lost sight of is that it is quite generally the case that what is conveyed properly includes and therefore entails what is said. (The only exceptions are cases of irony, metaphor, hyperbole and any other cases in which what is said is literally false and what is conveyed is therefore something other than what is said.) So we can now derive the maximally general and maximally useless scheme for all non-quality-based implicatures shown in (36).

    36. Non-Quality-Based Implicata: S entails W
                                         "W" conveys S
                                         "W" implicates S minus W

     The scheme in (36), which "predicts" implicatures only by subtracting the semantics of a sentence from what it is observed to convey when uttered, would seem to define the ultimate in non-explanation. Unfortunately, however, the Horn, Atlas and Levinson account will be just as utterly unexplanatory until some mechanism is proposed for predicting which entailing propositions are relevant for predicting the implicatures of an uttered proposition and which are not.
     When one considers this last fact in the light of all the other criticisms of radical pragmatics developed in this paper, one finds oneself at a loss for reasons to believe that radical pragmatics is even on the right track. Instead, it would seem high time to acknowledge the undeniable: no theory that puts as high a premium on simplicity and generality as does radical pragmatics will have any hope of accounting for the complexities invariably uncovered when pragmatic phenomena are examined in detail. Since radical pragmaticians themselves have produced most of the detailed examinations on which we base this conclusion, we have great hope that the alluringly bold but ultimately futile promises of radical pragmatics will soon begin to lose their appeal to those who honestly value full descriptive adequacy.

References

Atlas, J. and S. Levinson. 1981. It-clefts, informativeness and logical form. In Cole (ed.), 1981, 1-61.

Cole, P. ed. 1978. Syntax and Semantics 9: Pragmatics. New York: Academic Press.

---. ed. 1981. Radical pragmatics. New York: Academic Press.

Eilfort, W., P. Kroeber and K. Peterson. eds. Papers from the twenty-first regional meeting of the Chicago Linguistic Society. Chicago: Chicago Linguistic Society.

Farley, A., P. Farley and K.-E. McCullough. eds. 1986. Papers from the parasession on pragmatics and grammatical theory. Chicago: Chicago Linguistic Society.

Goldsmith, J. 1985. A principled exception to the coordinate structure constraint. In Eilfort *et al* (eds.), 133-43.

Grice, H.P. 1989. Studies in the ways of words. Cambridge: Harvard University Press.

Horn, L. 1972. On the semantic properties of logical operators in English. Bloomington: Indiana University Linguistics Club.

---. 1978. Remarks on neg-raising. In Cole (ed.), 1978, 129-220.

---. 1984. Toward a new taxonomy for pragmatic inference: q-based and r-based implicature. In Schiffrin (ed.), 11-42.

---. 1989. A natural history of negation. Chicago: University of Chicago Press.

Kempson, R. 1986. Ambiguity and the semantics-pragmatics distinction. In Travis (ed.), 77-103.

Lakoff, G. 1986. Frame semantic control of the coordinate structure constraint. In Farley *et al* (eds.), 152-67.

---. 1987. Women, fire and dangerous things. Chicago: University of Chicago Press.

Levinson, S. 1987. Minimization and conversational inference. In Papi et al (eds.), 61-129.

Morgan, J. 1978. Two types of convention in indirect speech acts. In Cole (ed.), 1978, 261-80.

Papi, M. and J. Verschueren. eds. 1986. The pragmatic perspective. Amsterdam: Benjamins.

Prince, E. 1982. Grice and universality: a reappraisal. University of Pennsylvania, MS.

Sadock, J. 1978. On testing for conversational implicature. In Cole (ed.), 1978, 281-97.

---. 1984a. Whither radical pragmatics? In Schiffrin (ed.), 139 - 49.

---. 1984b. The poly-redundant lexicon. In Testen *et al*, 250-69.

Schiffrin, D. ed. 1984. Meaning, form and use in context: linguistic applications. Georgetown University Roundtable 84. Washington: Georgetown University Press.

Sperber, D. and D. Wilson. 1986. Relevance. Cambridge: Harvard University Press.

Testen, D., V. Mishra and J. Drogo. 1984. Papers from the parasession on lexical semantics. Chicago: Chicago Linguistic Society.

Travis, C. ed. 1986. Meaning and interpretation. Oxford: Blackwell.

INDIRECTNESS IN CONVERSATION:
FLOUTING GRICE'S MAXIMS AT DINNER*

Suellen Rundquist
University of Minnesota

1. Introduction: Gender-differentiated language. There is a popular belief in our society that women's speech is more indirect than men's speech. Consistent with this belief, Lakoff 1975 argues that women's speech is more polite than men's speech, for example that women use more tag questions, hedges and mitigating language than men, and she proposes that women tend to speak not only more politely, but also more indirectly than men. However, a number of empirical studies have shown a lack of correlation between sex of the speaker and politeness (cf. Dubois and Crouch 1975, Baroni and D'Urso 1984, Brouwer, Gerritsen, and De Hann 1979, and Brouwer 1982). While these studies cast doubt on the claim that women are more polite, they do not question the premise that women are more indirect.[1]

There have been several studies dealing with gender-differentiated language and indirectness within the field of child language acquisition. Bellinger and Gleason 1982 (see also Gleason and Greif 1983) approached the topic of indirectness by studying mothers' and fathers' use of directives when talking to children in a laboratory setting. They isolated three types of directives: (1) imperatives, such as 'Give me the screwdriver', which were the most direct; (2) indirect questions, such as, 'Would you give me the screwdriver?'; and (3) implied directives, like 'I could really use the screwdriver', which were the least direct. They found that mothers used more indirect questions, which are considered to be the more polite forms, while fathers used both imperatives, the most direct, and implied directives, which are the least direct. So both fathers and mothers were indirect, but preferred to use different forms of indirectness.

Here I present a study which explores the relationship between indirectness and gender of the speaker, examining the speech of parents when talking to their children.

2. Conversational implicature and the cooperative principle. Grice 1975 examines the notion of indirectness, distinguishing CONVENTIONAL IMPLICATURE, where the conventional meaning associated with words or structures used in an utterance will determine what is implicated, from CONVERSATIONAL IMPLICATURE, which is what is implied or suggested beyond the conventional meaning, within a specific context in conversation.

As the basis for conversational implicature, Grice proposed the Cooperative Principle and maxims of conversation, as given in (1).

(1) Cooperative Principle (CP): Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.
   1. Maxim of Quantity (Do not say more or less than is required.)
   2. Maxim of Quality (Do not say what you believe to be false or that for which you lack evidence.)
   3. Maxim of Relation (Be relevant.)
   4. Maxim of Manner (Avoid obscurity, ambiguity; be brief and orderly.)

When the maxims are adhered to in conversation, the speaker intends to convey the content of an utterance as well as inferences beyond the semantic content of the utterance. Grice's theory has a social basis, where via the maxims, speakers learn from childhood to use indirectness in conversation and to interpret the meanings intended in indirect speech.

3. Flouting. Grice also discussed four ways in which it is possible for a participant to FAIL TO FULFILL a maxim. First, the speaker may VIOLATE a maxim by deliberately misleading the hearer, for example by telling a lie. Second, s/he may OPT OUT, making it clear that s/he is not willing to cooperate, e.g. 'my lips are sealed'. Third, there are situations in which TWO OR MORE MAXIMS MAY CLASH, where fulfilling one may violate another. And fourth, the speaker may OVERTLY NOT OBSERVE a particular maxim. Overtly not observing a maxim, which is referred to as FLOUTING, is a particular type of indirectness. When flouting a maxim, the speaker deliberately does not observe the maxim, and s/he expects the addressee to know that the maxim is not being observed; thus the Cooperative Principle is still in effect. According to Grice, irony, metaphor, meiosis and hyperbole are all instances of the speaker flouting a maxim, specifically the maxim of quality, i.e., do not say what you believe to be false. Grice provides a number of examples of flouting, one of which I present here as number (2), flouting the maxim of quantity by not giving enough information in a letter of recommendation:

(2)   Dear Sir, Mr. X's command of English is excellent, and his attendance at tutorials has been regular.

4. The study. In order to investigate the claim that women are more indirect than men, I decided to study one type of indirectness, namely what Grice refers to as flouting, in natural conversation, comparing and contrasting its use in men's and women's speech. Specifically, I am examining men's and women's speech in the following situations: (1) mothers and fathers speaking to their children, and (2) male and female participants in informal conversation, where no children are present. In this paper I discuss the results of a preliminary study of the first situation, with data collected from three families.

Based on my own observations of men interacting with their children, and hearing with some frequency exchanges such as the one illustrated in (3), I have suspected that men are indeed often indirect in speech.

(3)   (The child walks into the kitchen and takes a handful of popcorn.)
         Father: I thought you were practicing your violin.
         Child: I need to get the stand.
         Father: Is it under the popcorn?

The father's first statement in (3) is indirect and generates the conversational implicature that the child should be practicing. It is in fact an indirect directive to practice. The father's second utterance is even more indirect, and, since the question is not sincere, here the father is flouting the maxim of quality. I have thus formulated the hypothesis that when speaking to children, men flout Grice's maxims more than women do. In order to test this hypothesis for the pilot study, I asked three families to tape four evenings of dinner conversation, all with children present. In analyzing the data, I specifically addressed the following questions: How often did the father and mother flout the maxims? Did the father in fact flout the maxims more often than the mother did, and when flouting occurred, who was the addressee? Are some maxims flouted more often than others? And finally, Why do people flout the maxims? What functions does flouting perform?

4.1. The participants. The people participating in the study are friends and neighbors with whom I am quite well acquainted, including members of my own family. In order to control variables such as education, age, and socio-economic level for the preliminary study I chose families who are similar in many respects. All parents are

college graduates between the ages of 35 and 40; both parents in each family are professionals, with all mothers working outside the home. Each family has two children ranging in age from five to eleven, all of them well beyond the initial stages of language acquisition, but not yet adolescents or young adults. A disadvantage of studying such a homogeneous group is that the conclusions reached concerning these data cannot be generalized to the entire population, but as Jenkins 1985:147 notes, 'contrary to traditional social science dictum, you can sometimes learn more by first observing and interviewing people you know well, rather than those you don't'; similarly, Tannen 1984:33 justifies recording the conversation of friends at dinner by saying that 'recording a conversation among friends that would have taken place anyway makes available for study patterns of language use that do not emerge among strangers, such as playful routines, irony and allusion, reference to familiar jokes and assumptions.'

    4.2. Methods of collecting the data. Each family has agreed to record four evenings of dinner table conversation within a reasonable span of time.[2] As stated previously, what is presented here is from the initial study of three families.

    The method of collecting the data involves asking the participants to turn on a taperecorder at meal times, and to turn it off when they finish the meal. The assumption is that they will forget about the taperecorder as they become involved in their day-to-day conversation. Fishman 1983 studied the conversation of three couples in their own homes. The participants agreed to allow a taperecorder to be placed in their apartments, but they also had censor rights and could turn it on and off at will. 'All six people reported that they soon began to ignore the tape recorder' (p. 92). In response to the possibility that participants are aware of a taperecorder and thus might not speak naturally, Tannen states that 'if there is a relatively large number of participants who have ongoing social relationships, they soon forget the taperecorder' (p. 34). In my study, initially the taperecorder was a distraction for each family. Children in particular 'performed' at the beginning of each meal. However, as the meal progressed the recorder was forgotten and conversation became more spontaneous.[3]

    4.3. General results. The results of this study show that, overall, the fathers in each family do flout the maxims more often than the mothers do. In the tables that follow the participants in this study will be known as M1, M2, M3, where M=Mom, and D1, D2, D3, where D=Dad.

	Total number of utterances	Number of flouted utterances	% of flouted utterances	Mean
D1	395	92	23.30	
D2	267	52	19.50	
D3	256	16	06.25	16.35%
M1	468	34	07.30	
M2	330	19	05.75	
M3	267	05	01.90	04.98%

TABLE (1) FLOUTING THE MAXIMS AT DINNER: GENERAL RESULTS

Based on the total number of utterances for each speaker, as shown on Table (1), D1 flouted on the average of 23.3% of the time, whereas M1 flouted the maxims only 7.3% of the time. In the second family, D2 flouted the maxims 19.5% of the time, while M2 flouted only 5.75%; and in the third family, while D3 only flouted 6.25% of the time, M3 flouted even less, 1.9% of her total utterances. The percentages for individuals vary greatly, especially when comparing the results for D1 and D3. And even though D3

flouted much less than either D1 or D2, he still flouted much more than M3. Before the study was begun, my intuition was that both D1 and D2 are prone to communicating by flouting. The results indicate that flouting the maxims is more a part of the conversational habits of these two men than it is with D3. Based on these data, I cannot say that men in general are prone to flouting the maxims, but it seems clear that these fathers all flout the maxims more than the mothers do. It should be noted that in all three families the father flouted over three times as often as the mother did.

Participant	Child as Addressee	Adult as Addressee	Other
D1	78%	22%	2%
M1	59%	41%	0
D2	79%	15%	6%
M2	68%	32%	0
D3	62.5%	37.5%	0
M3	60%	40%	0

TABLE (2) FLOUTED UTTERANCES ACCORDING TO ADDRESSEE

As shown on Table (2), all participants flouted the maxims more often when the addressee was a child than when the addressee was an adult. In each family, however, the percentage of flouts to children is greater for fathers than for mothers (although in Family 3 the difference is negligible).

Because the study involved two adults and two children at the dinner table, when a speaker addressed an utterance to one person, the other two people, though not addressed, heard what was said. Clark and Carlson 1982 propose four basic roles in conversations: the SPEAKER, who then determines who fills the other roles; the PARTICIPANTS, those meant to take part in the conversation; the ADDRESSEE(S), directly spoken to and also included as participant(s); and the OVERHEARERS, those listening but not intended to take part in the conversation. Relating this to dinner conversation in Family 1, when D1 addressed M1, the children were still participants in the conversation. When D1 addressed an utterance to M1 which flouted a maxim, the children, as participants, were also meant to hear the utterance. One could speculate that even though 78% of D1's flouts were directly addressed to a child, 100% of the examples of flouting could be due to having the children as participants.

The above situation exists in Families 2 and 3 as well. Even though a number of flouted utterances were directed to an adult, the children were in fact participants in the conversation.[4]

I have only a small amount of data without children present, two evenings of conversation in Family 1 after the children had left the table, and there is no evidence of flouting by either D1 or M1. Also, one evening after one child had left the table, the other child was treated as what Clark and Carlson refer to as an overhearer rather than a participant; she was present but not intended to take part in the conversation. The content was such that it did not involve her and she was not interested. No flouting by either M1 or D1 took place during this part of the conversation. These data suggest that both mothers and fathers flout the maxims more frequently when conversing with children as participants than they do when in adult conversation, an hypothesis that will be tested later in the study.

4.4. Maxims flouted. Grice states that 'the presence of a conversational implicature must be capable of being worked out' as follows: 'To work out that a particular conversational implicature is present, the hearer will reply on the following data:

(1) the conventional meaning of the words used, together with the identity of any references that may be involved; (2) the CP and its maxims; (3) the context, linguistic or otherwise, of the utterance; (4) other items of background knowledge; and (5) the fact (or supposed fact) that all relevant items falling under the previous headings are available to both participants and both participants know or assume this to be the case' (p. 50). Under these circumstances I have found that it is possible on occasion to arrive at more than one interpretation of an utterance. I have also found that any given utterance may in fact flout more than one maxim. All possibilities found have been included in the percentages on Table (3).

	D1	M1	D2	M2	D3	M3
Quality	50%	50%	75%	83%	75%	75%
Quantity	26%	31%	08%	33%	06%	25%
Manner	16%	14%	15%	08%	06%	25%
Relation	08%	03%	02%	0	13%	0

<center>TABLE (3) MAXIMS FLOUTED</center>

Note that, for all adult participants, the maxim of quality, do not say what you believe to be false, was the most often flouted maxim, whereas the maxim of relation or relevance was flouted least often for five of the six participants.

4.5. Examples and discussion. In this section I present examples from the data, after which I discuss possible interpretations. First of all, typical of at least one conversation in each family and several in two of the families, was the situation where the father would flout on a given topic, and both he and the mother would continue flouting on that topic through a portion of the conversation, as shown in examples (4a) through (4d).[5]

(4a)   (The child is not eating his dinner. The child hates corn, but corn is NOT being served at this meal. There is no corn in the house.)
        D[12]:   You eat what you have there, you don't have to eat your corn.

(4b)   D[33]:   Say, ____, am I going to have to go buy corn?
        Child: No.

(4c)   D[36]:   You have to eat your dinner. Otherwise It's gonna be corn and Bert [a troll from a story] will sit on you and squish you all to jelly if you don't eat your corn.
        Child: [In exaggerated tone] Oh no!

(4d)   M[45]:   You have to finish everything or I'll start making corn.

In contrast, none of the flouts initiated by mothers were continued or built upon by fathers. In other words the father was the instigator of flouting, whereas the mother went along with it. One possible conclusion this leads to is that since many of the mothers' flouts were prompted by the fathers' flouts, mothers might have flouted even less than they did had the fathers not been part of the conversation. Thus the fathers appear to play a dominant role in this type of speech. Tannen 1984 noted a similar trend when discussing humor in her conversation data. A male participant was the initiator of humorous remarks and a female participant built on them.

There are several examples in the data of men expressing superficial self-criticism, as in example (5).

(5)   (Discussion topic is television programs. Child has been explaining what happened on 3-2-1 Contact.)
      D: I don't think I ever watched 3-2-1 Contact. I never got beyond Electric Company. I thought Electric Company was kinda hard.

The father is not being sincere in his comment, and everyone knows this. In the conversations in this study neither the children nor the mothers engaged in this behavior.

Parents frequently play a role in regulating the behavior of their children. In the families studied here there is evidence of gender differences in how this regulation of behavior is expressed. Both parents asserted their authority, both directly and indirectly, but fathers expressed it by flouting more often than mothers did.

An obvious way that parents regulate behavior is by use of directives. The fathers in this study, much more often than the mothers, performed directives by flouting, for example those illustrated in (3) and (4) above. Note in (4a) and (4c) that the father starts with a directive, but immediately tones it down by flouting, whereas in (4b) the entire utterance is an example of flouting. The exchange in (6) shows a nice contrast between the approach of mother and father:

(6)   (Child is not eating and complains that some of the noodles on his plate are not cooked.)
      M: Well, then don't eat them.
      D: The banana's not cooked either.

Whereas the mother is more direct, the father is rather obscure, flouting the maxim of relevance as well as that of manner. The immediate implication is that the child eats bananas even though they are not cooked. Based on his previous remarks it is clear that he thinks the child SHOULD eat everything.

Another way that parents regulate their children's behavior is by criticism. Again fathers criticized by flouting more than mothers did, as illustrated in (7) and (8).

(7)   M: There's one piece [of pie] left.
      Child: What is it? Rhubarb? Oh.
      D: We can split it.
      Child What's it made of?
      M: Rhubarb.
      D: Pineapple.
      Child: That's original!

In example (7), first of all note again the contrast in parental approach. The mother's response is direct, while the father's response generates the implicature that the child's question was inappropriate or silly, since pineapple is not an ingredient in rhubarb pie, and the father and child both know this.[6] The child's retort is actually parroting a comment that the father made earlier in the conversation.

(8)   (Child has been refusing to speak because she is angry with her father. Now she is making exaggerated motions to indicate that she is having trouble cutting her piece of pizza.)
      D: Would you like a hatchet?

In example (8), the father is not sincere, since a hatchet is an unlikely tool for a child to use on a piece of pizza, implicating that the child is making too much fuss over her pizza-cutting problem, but also that he thinks that she should be talking.

Although the above examples show that fathers flout when attempting to regulate their children's behavior, in a number of utterances this position of power is softened by the use of flouting. In the following situation the child is still angry with her dad for shouting at her. She refuses to speak throughout a portion of the conversation. Finally, she begins to make hand motions as if she wants to contribute.

(9)   D: Sounds like...one syllable...first word.
(Child continues to make undecipherable motions.)
         D: One...One...let's see...IXI...so that's...that's ten!

In (9) by being silly and teasing the child, the father is actually trying to ease her back into the conversation. He is successful, as several utterances later she is again an active speaker. She asks where the oars and pump for a rubber raft have come from, and the father responds with (10).

(10)   D: The oars and the pump were purchased separately by your generous father...masquerading as the mean ogre...who lives here...and shouts.
         Child: And doesn't like people to talk!

The father's utterance here is mitigating in nature, an apology of sorts, but note that the child, even though she is now speaking, hasn't quite forgiven him.

There are several hypotheses available through which one can analyze these examples. Most prevalent in the literature on gender-differentiated speech is the notion of power and dominance (for example see Fishman 1983, West and Zimmerman 1983). As shown in example (4) it seems clear that fathers play a dominant role in flouting. They use it in directives as well as in criticism when speaking to these children. People in a position of power can be confident in expressing their dominance through humor and teasing, or more severely through sarcasm. In analyzing the data I asked the child involved how she felt about the remarks in (7) and (8). She remembered the instance in (7) as being funny, but in looking at (8) she remembered being irritated. She indicated that what is doubly maddening about remarks like this is that they are funny as well as being irritating. When I asked her if she thought her dad was being mean when he talked like this, she responded, 'No, he can't help it. That's just the way he talks.' One final point in support of the power factor is that people who are older and bigger can also afford to put themselves down, being secure in the fact that the addressee will know that it is not to be taken seriously. This is illustrated in (5), where the father says that he always thought Electric Company was too hard. It appears that he is implicating that he is in fact bigger and more knowledgeable than the children are.

Another possible hypothesis (cf. Maltz and Borker 1982) is that men prefer to show themselves off to their best advantage in conversation. They tend to posture, to want a dominant role in the conversation, to want to be center-stage. Many of the examples in my data are humorous. The participants in the conversation do laugh. All of these examples could be interpreted as men's attention-getting strategies. When these men flout the maxims they continuously reassert their control over the conversation by getting everyone's attention. To support the interpretation that these men are trying to be entertaining, there are several instances of children saying, 'Dad, you're crazy,' or 'Don't be dumb,' after the father has flouted a maxim. The children do not appear to be threatened by flouting when they can respond in this way.

Finally, although there is evidence that men prefer direct confrontation, it has been suggested that men, when interacting with women in some situations, prefer to avoid

direct confrontation, also avoiding emotional scenes. Example (4), where the father continuously tries to avoid a direct command to the child to eat his dinner, could be interpreted as avoidance of conflict with the child. If the father had said, 'Eat your dinner,' the child could more easily reply with a negative. What in fact happens here is that the child picks up on the idea of the illusive corn. In (4b) he responds 'no' to his father's having to go buy corn, not to the directive that he should be eating his dinner; whereas in (4c) he responds in mock horror to the idea of being squashed by a troll. In (7), the rhubarb/pineapple example, if the father had said, 'That's a stupid question,' instead of flouting a maxim, the child would have been more likely to take offense, be hurt, and possibly even cry. As it is, she plays his game, responding as he might respond in a similar situation. The examples in (9) and (10) of appeasing and mitigating behavior fit well into this theory, as avoidance of a directly expressed apology.

Gleason and Greif 1983, in examining studies of parents and children, note that fathers tend to be 'impolite or insensitive to their children during conversation' (p. 147), but that they also have a style that demands more from children, e.g. they use more difficult vocabulary, request more sophisticated information, and encourage the children to display their knowledge. The authors propose that the father's use of implied directives requires more inferencing from the child. They conclude that mothers are more polite and sensitive to the child's conversation, more 'tuned-in' to the child's communication, which means that the child does not have to communicate well verbally for the mother to understand; whereas the father, demanding more precise communication skills, serves as a 'bridge' to the community, where the child will have to express her/himself clearly in order to be understood. Their hypothesis provides another perspective in examining the speech of the fathers in this study. Since flouting requires more inferencing than a direct statement, one could speculate that the frequent use of flouting challenges the child's ability to inference. The reasons the fathers flout could be any of the above mentioned possibilities, a power ploy, an attention getter, or an avoidance of direct confrontation, but one effect on the children might be that, through exposure to this type of speech at home, they are better able to deal with inferencing in general.

5. Conclusions and implications. Based on these data, the preliminary hypothesis that, when speaking to children, men flout Grice's maxims more than women do, is upheld. Thus, with respect to at least one type of indirectness, flouting the maxims, men's speech appears to be more indirect than women's speech.

Since in the small amount of data presented here without children present no flouting occurred, I propose several questions for further research: Do men flout the maxims more than women do when children are not participants in the conversation? Does flouting by either men or women occur at the same frequency when children are not present as when children are present? In other words, does the presence of children increase or decrease the amount of flouting in natural conversation? And, does the sex of the child play a role in frequency of flouting by either parent? Two of the three families in the pilot study have a boy and a girl, and each parent flouted to both more or less equally; but whether the sex of the child plays a role in frequency of flouting is an important question to continue to address in the larger study. In order to answer these questions, my study continues on two fronts. First, more families have agreed to participate, thus increasing the amount of data; and second, all parents involved have agreed to be recorded in conversation at a social occasion without children.

In conclusion, Sperber and Wilson 1986 propose an alternative to Grice's theory of indirectness. Their theory of relevance puts all types of speech on a continuum from most explicit to least explicit, deriving all inferences from a single principle of relevance: thus the relevance of the utterance to the situation is the only factor involved. Their theory also differs from Grice's in that it has a cognitive rather than a social basis. Within this theory one would be less likely to expect social variables to interact with degrees of indirectness. Ziv 1988:539 has noted that Sperber and Wilson's approach cannot explain

social conventions such as politeness. 'For politeness to be subsumed under the PRL [Principle of Relevance] it would have to be considered an aspect of cognition, which it is clearly not. The essence of politeness is social, and as such the PRL cannot account for it.' The results of my study show tentatively that there are differences in how men and women use implicature, i.e. that the social variable of gender is a significant factor in the use of indirectness. This also seems to provide evidence for a theory with a social component as well as a cognitive component. Thus my study not only bears on the specific question of gender differences and indirectness, but also on the broader issue of whether conversational inference is based entirely on a cognitive foundation.

[1]But see Tannen 1982 on interpreting indirectness in male-female discourse.

[2]The first family recorded six sessions, but I asked the others to do only four.

[3]One adult (D2), though, did speak to the taperecorder and occasionally treated me as a participant in the conversation.

[4]D1 and D2 both have a small number of utterances directed towards no particular addressee or towards a pet. It may be the presence of children as participants that instigated this behavior.

[5]The numbers given after D or M indicate the utterance number of that person, illustrating the fact that these utterances were distributed over a portion of the conversation.

[6]It has been suggested to me that the mother's response is also a flout, that she is flouting the maxim of quantity by giving too much information.

REFERENCES

Baroni, Maria Rosa, and Valentina D'Urso. 1984. Some experimental findings about the question of politeness and women's speech (Research note). Language in Society 13.67-72.

Bellinger, David C., and Jean Berko Gleason. 1982. Sex differences in parental directives to young children. Sex Roles 8.1123-39.

Bremner, Sue, Noelle Caskey, and Birch Moonwomon (eds.). 1985. Proceedings of the First Berkeley Women and Language Conference. University of California, Berkeley.

Brouwer, D. 1982. The influence of the addressee's sex on politeness in language use. Linguistics 20.697-711.

_____, Marinel Gerritsen, and Dorian De Haan. 1979. Speech differences between women and men: on the wrong track? Language in Society 8.33-50.

Clark, Herbert H., and Thomas B. Carlson. 1982. Hearers and speech acts. Language 58.332-373.

Cole, Peter, and Jerry L. Morgan (eds.). 1975. Syntax and semantics: speech acts, vol. 3. New York: Academic Press.

Dubois, Betty Lou, and Isabel Crouch. 1975. The question of tag questions in women's speech: they don't really use more of them, do they? Language in Society 4.289-294.

Fishman, Pamela M. 1983. Interaction: the work women do. In Thorne, Kramarae, and Henley, 89-101.

Gleason, Jean Berko, and Esther Blank Greif. 1983. Men's speech to young children. In Thorne, Kramarae, and Henley, 140-149.

Grice, H. P. 1975. Logic and conversation. In Cole and Morgan, 41-58.

Gumperz, J. J. (ed.). 1982. Language and social identity. Cambridge University Press.

Jenkins, Mercilee M. 1985. What's so funny?: joking among women. In Bremner, Caskey, and Moonwomon, 135-151.

Lakoff, Robin. 1975. Language and woman's place. New York: Harper and Row.

Maltz, Daniel N., and Ruth A. Borker. 1982. A cultural approach to male-female miscommunication. In Gumperz, 196-216.

Sperber, Dan, and Deirdre Wilson. 1986. Relevance: communication and cognition. Cambridge, Massachusetts: Harvard University Press.

Tannen, Deborah. 1982. Ethnic style in male-female conversation. In Gumperz, 217-231.

_____. 1984. Conversational style: analyzing talk among friends. Norwood, New Jersey: Ablex Publishing Company.

Thorne, Barrie, Cheris Kramarae, and Nancy Henley. 1983. Language, gender and society. Rowley, Massachusetts: Newbury House Publishers, Inc.

West, Candace, and Don H. Zimmerman. 1983. Small insults: a study of interruptions in cross-sex conversations between unacquainted persons. In Thorne, Kramarae, and Henley, 103-117.

Ziv, Yael. 1988. On the rationality of 'relevance' and the relevance of 'rationality'. Journal of Pragmatics 12.535-545.

Rethinking Power and Solidarity
in Gender and Dominance

Deborah Tannen
Georgetown University

In analyzing discourse, many researchers assume that all speakers proceed along similar lines of interpretation. But Gumperz 1982 makes clear that this is so only to the extent that cultural background is shared. To the extent that cultural backgrounds differ, lines of interpretation are likely to diverge. My own research shows that cultural difference is not limited to the gross and apparent levels of country of origin and native language, but also exists at the subcultural levels of ethnic heritage, class, geographic region, age, and gender. My earlier work focused on ethnic and regional style; my most recent work focuses on gender. I draw on this work here to demonstrate that utterances have widely divergent potential implicatures, because of the ambiguity of power and solidarity.

Power and Solidarity. Since Brown and Gilman's 1960 pioneering study, and the subsequent contributions of Friedrich 1972 and Brown and Levinson [1978]1987, the concepts of power and solidarity have been fundamental to sociolinguistic theory. (Fasold 1990 provides an overview.) Power is associated with nonreciprocal forms of address: a speaker addresses another by title-last-name but is addressed by first name. Solidarity is associated with reciprocal forms of address: both speakers address each other by title-last-name or first name. Power governs asymmetrical relationships where one is subordinate to another; solidarity governs symmetrical relationships characterized by social equality and similarity. In my own work (Tannen 1984, 1986) exploring the relationship between power and solidarity as it emerges in conversational discourse, I argue that although power and solidarity, closeness and distance, seem at first to be opposites, they also entail each other. Any show of solidarity necessarily entails power, in that claiming similarity and closeness limits freedom and independence. At the same time, any show of power entails solidarity by involving participants in relation to each other.

I once entitled a lecture "The Paradox of Power and Solidarity." The respondent to my talk appeared wearing a three-piece suit and a knapsack on his back. On one level, the suit represented power, the knapsack solidarity. Now wearing a knapsack would also mark solidarity at, say, a protest demonstration. And wearing a three-piece suit to the demonstration might mark power by differentiating the wearer from the demonstrators, perhaps even reminding them of his superordinate position in an institutional hierarchy. But wearing a three-piece suit to the board meeting of a corporation would mark solidarity, and wearing a knapsack in that setting would connote not solidarity but disrespect, a move in the power dynamic.

This ambiguity gives rise to what has been a major thrust in my previous decade's work analyzing conversation: What appear as attempts to dominate a conversation (an exercise of power) may actually be intended to establish rapport (an exercise of solidarity). This occurs because power and solidarity are bought with the same currency: The same linguistic means can be used to create either or both. The paradigm example from my analysis of a two and a half hour dinner table conversation

(Tannen 1984) involves interruption, or, more accurately, overlap, which can be interpreted as interruption. For some, speaking along with another is a show of enthusiastic participation in the conversation (solidarity); for others, only one voice should be heard at a time, so any overlap is an interruption (power). The result was that enthusiastic listeners who overlapped cooperatively, talking along to establish rapport, were perceived as interrupting. This doubtless contributed to the impression reported by the other speakers that they had "dominated" the conversation. Indeed, the tape and transcript also give the impression that they had dominated, because their overlap-aversant friends tended to stop speaking as soon as another voice began.

The key to solving the puzzle of whether power or solidarity is primary is symmetry. If Speaker A repeatedly overlaps and Speaker B repeatedly gives way, the resulting communication is asymmetrical, and the effect is domination. But if both speakers avoid overlaps, or if both speakers overlap each other and win out equally, there is symmetry and no domination.

The dynamics of power and solidarity are also seen in the tension between similarity and difference. Scollon 1982:344-5 explains that all communication is a double bind because of the conflicting needs to be left alone (negative face) and to be accepted as a member of society (positive face). Becker 1982:125 expresses this double bind as "a matter of continual self-correction between exuberance (i.e. friendliness: you are like me) and deficiency (i.e. respect: you are not me)." I (Tannen 1984:17) have described this double bind as inherent in communication in that we send and receive the simultaneous and conflicting messages, "Don't assume I'm the same as you" and "Don't assume I'm different from you." All these formulations elaborate on the tension between similarity and difference, or what Becker and Oka 1974 call "the cline of person," a semantic dimension they suggest may be the one most basic to language; that is, one deals with the world and the objects and people in it in terms of how similar, or close, they are to oneself.

In my most recent work (Tannen 1990c) I have continued to explore the paradoxical and mutually entailing relationship between power and solidarity by examining how conflict is negotiated in conversational interaction. This involves exploring how what appears as dominance may not be intended as such, but also how what seems like cooperation may actually be intended to dominate. Throughout, I demonstrate that the linguistic strategies by which power and solidarity are achieved and expressed can be the same, so intentions such as dominance cannot be correlated with linguistic strategies. Rather, the "meaning" of any strategy depends on context, the conversational styles of participants, and the interaction of their styles and strategies with each other.

Similarity and Difference. Harold Pinter's most recent play <u>Mountain Language</u>, composed of four brief scenes, is set in a prison in the capital city of an unnamed country. In the second scene, an old mountain woman is finally allowed to see her son across a table as a guard stands over them. But whenever she tries to speak to her son, the guard silences her, telling the prisoner to tell his mother that their mountain language is forbidden. Then he continues:

GUARD: And I'll tell you another thing. I've got a wife and three kids. And you're a pile of shit.
Silence.
PRISONER: I've got a wife and three kids.
GUARD: You've what? Silence. You've got what?
Silence. What did you say to me? You've got what? Silence. You've got <u>what?</u>
[He picks up the telephone and dials one digit.] Sergeant? I'm in the Blue Room
... yes ... I thought I should report, Sergeant ... I think I've got a joker in here.

The Sergeant soon enters and asks, "What joker?" The stage darkens and the scene ends. The final scene opens on the same setting, with the prisoner bloody and shaking, his mother shocked into speechlessness. The prisoner was beaten for saying, "I've got a wife and three kids." This quotidian statement, which would be unremarkable in casual conversation, was insubordinate in the hierarchical and oppressive context because the guard had just made the same statement. When the guard said, "I've got a wife and three kids. And you're a pile of shit," he was claiming, "I am different from you." By repeating the guard's words verbatim, the prisoner was saying, "I am the same as you." (I have demonstrated at length [Tannen 1987, 1989a] that repeating another's words creates rapport on a meta level.) The guard was asserting his own humanity and denying the prisoner's; by claiming <u>his</u> humanity and implicitly denying the guard's assertion that he is "a pile of shit," the prisoner challenged the guard's right to dominate him. Similarity is antithetical to hierarchy.[1]

The ambiguity of closeness, a spatial metaphor representing similarity or involvement, emerges in a nonverbal aspect of this scene. In the performance I saw, the guard repeated the question "You've got what?" while moving steadily closer to the prisoner, until he was bending over him, nose to nose. The guard's moving closer is a nonverbal analogue to the prisoner's statement, but with opposite effect: he was "closing in." The guard moved closer and brought his face into contact with the prisoner's not as a sign of affection (which such actions could signify in another context) but as a threat. Closeness, then, can mean aggression rather than affiliation in the context of a hierarchical rather than symmetrical relationship.

<u>The Ambiguity of Linguistic Strategies</u>. The potential ambiguity of linguistic strategies to mark both power and solidarity in face-to-face interaction has made mischief in language and gender research, wherein it is tempting to assume that whatever women do results from their powerlessness. But all the linguistic strategies that have been taken by analysts to be evidence of dominance can under certain circumstances be instruments of affiliation. These include indirectness, interruption, silence vs. volubility, topic raising, and conflict and verbal aggression. For the remainder of this paper, I demonstrate the ambiguity of each of these strategies in turn.

<u>Indirectness</u>. Lakoff 1975 identifies two benefits of indirectness: defensiveness and rapport. Defensively, a speaker can later disclaim or rescind an indirect communication that does not meet with a positive response. At the same time, indirectness yields the pleasant experience of getting one's way not because one demanded it (power) but because the other wanted the same thing (solidarity). Many researchers have focused on the defensive or power benefit of indirectness and ignored the payoff in rapport or

solidarity. Conley, O'Barr and Lind's 1979 claim that indirectness is the language of the powerless has been particularly influential. Women are then seen as indirect because they don't feel entitled to make demands. Yet those who feel entitled to make demands may prefer not to, seeking the payoff in rapport. Furthermore, there are circumstances in which a party is so powerful that it is unnecessary to make a demand. For example, an employer who says, "It's cold in here" may expect a servant to close a window, but a servant who says the same thing is not likely to see the employer rise to correct the situation. This may explain the otherwise surprising finding of Bellinger and Gleason 1982 (reported in Gleason 1987) that fathers' speech to their young children had a higher incidence than mothers' of both direct imperatives ("Turn the bolt with the wrench") and implied indirect imperatives ("The wheel is going to fall off").

Cultural relativity sheds crucial light on the use of indirectness. Keenan 1974 found that in a Malagasy-speaking village on the island of Madagascar, women are direct and men indirect, although the men are socially dominant and their indirect style more highly valued. My own research (Tannen 1981) found that whereas American women were more likely to take an indirect interpretation of a sample conversation, Greek men were as likely as Greek women, and more likely than American men or women, to take the indirect interpretation. Indirectness, then, is not in itself a strategy of subordination.

Interruption. That interruption is a sign of dominance has been as widespread an assumption in research (for example, Leet-Pellegrini 1980) as in conventional wisdom. A frequent finding (for example, West and Zimmerman 1983) is that men dominate women by interrupting them in conversation. Tellingly, however, Deborah James and Janice Drakich (personal communication), reviewing research on gender and interruption, discovered that studies comparing amount of interruption in all-female vs. all-male conversations find more interruption, not less, in the all-female ones. Though initially surprising, this finding reinforces the need to distinguish linguistic strategies by their interactional purpose. Does the overlap show support for the speaker, or does it contradict or change the topic? Elsewhere (Tannen 1989b, 1990c) I explore at length the problems inherent in the claim that interruption can be identified by mechanical means, and I give examples of conversations in which there is overlap but no interruption and interruption but no overlap.

This is not, however, to say that interruption never constitutes dominance nor that men never interrupt or dominate women. Fictional discourse provides an example of a situation in which it is and one does. In a short story by Lorrie Moore, Zoe is talking to a man she has just met at a party. He asks, "What's your favorite joke?" When she begins, "A man goes to a doctor," he interrupts: "I think I know this one. A guy goes into a doctor's office, and the doctor tells him he's got some good news and some bad news -- that one, right?" It is obvious that this is not right, because Zoe's joke is "about the guy who visits his doctor and the doctor says, 'Well, I'm sorry to say, you've got six weeks to live.'" But instead of saying "No," Zoe says, "I'm not sure. This might be a different version," leaving open the door for him to tell his joke, which turns out to be not only different but offensively obscene. This interruption does seem dominating because it comes as Zoe is about to tell a joke and usurps the floor to tell it.

The point, then, is that in order to understand the "meaning" of an interruption, or, indeed, whether an overlap is an interruption, one must consider the context, speakers' styles, and the interactive frame -- that is, what the speakers are trying to do by their communication.

Silence vs. Volubility. The excerpt from Pinter's <u>Mountain Language</u> dramatizes the assumption that powerful people do the talking and powerless people are silenced. This is the trope that underlies the play's title and its central theme: By outlawing their language, the oppressors silence the mountain people, rob them of their ability to speak and hence of their humanity. In the same spirit, many scholars (for example, Spender 1980) have claimed that men dominate women by silencing them. Again, there are surely circumstances in which this is accurate. Coates 1986 notes numerous proverbs that instruct women, like children, to be silent.

Silence in itself, however, is not a sign of powerlessness, nor volubility a self-evident sign of domination. A theme running through Komarovsky's 1962 classic study <u>Blue Collar Marriage</u> is that many of the wives interviewed said they talked more than their husbands: "He's tongue-tied," one woman said; "My husband has a great habit of not talking," said another; "He doesn't say much but he means what he says and the children mind him," said a third. Yet there is no question but that these husbands are dominant in their marriages.

Indeed taciturnity itself can be an instrument of power. This is precisely the claim of Sattel 1983 who argues that men use silence to exercise power over women. He illustrates with a scene from Erica Jong's novel <u>Fear of Flying</u>, only a brief part of which is presented here. The first line of dialogue is spoken by Isadora, the second by her husband, Bennett. (Spaced dots indicate omitted text; unspaced dots show a pause included in the original text.)

> "Why do you turn on me? What did I do?"
> Silence.
> "What did I do?"
> He looks at her as if her not knowing were another injury.
> "Look, let's just go to sleep now. Let's just forget it."
> "Forget what?"
> He says nothing.
>   . . .
> "It was something in the movie, wasn't it?"
> "What, in the movie?"
> ". . . It was the funeral scene. ... The little boy looking at his dead mother. Something got you there. That was when you got depressed."
> Silence.
> "Well, <u>wasn't</u> it?"
> Silence.
> "Oh come on, Bennett, you're making me <u>furious</u>. Please tell me. Please."

The painful scene continues in this vein until Bennett tries to leave the room and Isadora tries to detain him. It certainly seems to support Sattel's claim that Bennett's silence subjugates his wife, as the scene ends with her literally lowered to the floor, clinging

subjugates his wife, as the scene ends with her literally lowered to the floor, clinging to his pajama leg. But the reason his silence is an effective weapon is her insistence that he tell her what's wrong. If she too receded into silence, his silence would be disarmed. The devastation results not from his silence in itself but from the interaction of their differing styles.

Volubility and taciturnity, too, can result from style differences rather than speakers' intention. As I (Tannen 1984, 1985) and others (Scollon and Scollon 1981, Scollon 1985) have discussed at length, there are cultural and subcultural differences in the length of pauses expected between and within speaking turns. In my study of dinner table conversation, those who expected shorter pauses between conversational turns began to feel an uncomfortable silence ensuing while their longer-pausing friends were simply waiting for what they regarded as the appropriate time to take a turn. The result was that the shorter pausers ended up doing most of the talking, another sign interpreted by their interlocutors as dominating the conversation. But their intentions had been to fill in what to them were potentially uncomfortable silences, that is, to grease the conversational wheels and ensure the success of the conversation. In their view, the taciturn participants were uncooperative, failing to do their part to maintain the conversation. So silence and volubility, too, may imply either power or solidarity.

Topic-raising. Shuy 1982 is typical in assuming that the speaker who raises the most topics is dominating a conversation. However, in a study I conducted (Tannen 1990a,b) of videotaped conversations among friends of varying ages recorded by Dorval 1990, it emerged that the speaker who raised the topics was not always dominant, as judged by other criteria (for example, who took the lead in addressing the investigator when he entered the room). To illustrate: in a twenty-minute conversation between a pair of sixth-grade girls who identified themselves as best friends, Shannon raised the topic of Julia's relationship with Mary by saying, "Too bad you and Mary are not good friends anymore." The conversation proceeded and continued to focus almost exclusively on Julia.

Similarly, most of the conversation between two tenth-grade girls was about Nancy, but Sally raised the topic of Nancy's problems. In response to Nancy's question "Well, what do you want to talk about?" Sally said, "Your mama. Did you talk to your mama?" Overall, Sally raised nine topics, Nancy seven. However, all but one of the topics Sally raised were questions focused on Nancy. If raising more topics is a sign of dominance, Sally controlled the conversation when she raised topics, although even this was subject to Nancy's collaboration by picking them up. It may or may not be the case that Sally controlled the conversation, but the nature of her dominance is surely other than what is normally assumed by that term if the topics she raised were all about Sally.

Finally, the effect of raising topics may also be an effect of differences in pacing and pausing, as discussed above with regard to my study of dinner table conversation. A speaker who thinks the other has no more to say on a given topic may try to contribute to the conversation by raising another one. But a speaker who was intending to say more and was simply waiting for the appropriate turn-exchange pause, will feel that the floor was taken away and the topic aggressively switched. Yet again, the impression of dominance might simply result from style differences.

Conflict and Verbal Aggression. Research on gender and language (see Maltz and Borker 1982 for a review) has consistently found male speakers to be competitive and more likely to engage in conflict (for example, by arguing, issuing commands, and taking opposing stands) and females to be cooperative and more likely to avoid conflict (for example, by agreeing, supporting, and making suggestions rather than commands.) Ong 1981:51 argues that "adversativeness" is universal, but "conspicuous or expressed adversativeness is a larger element in the lives of males than of females."

In my analysis of videotapes of male and female friends talking to each other (Tannen 1990b,c) I have begun to investigate how male adversativeness and female cooperation are played out, complicated, and contradicted in conversational discourse. In analyzing Dorval's videotapes of friends talking, for example, I found a sixth-grade boy saying to his best friend, "Seems like, if there's a fight, me and you are automatically in it. And everyone else wants to go against you and everything. It's hard to agree without someone saying something to you." In contrast, girls of the same age (and also of most other ages whose talk I examined) spent a great deal of time discussing the dangers of anger and contention. In affirming their own friendship, one girl told her friend, "Me and you <u>never</u> get in fights hardly," and "I mean like if I try to talk to you, you'll say, 'Talk to <u>me</u>!' And if you try to talk to me, I'll <u>talk</u> to you."

These examples of gendered styles of interaction are illuminated by the insight that power and solidarity are mutually evocative. As seen in the statement of the sixth grade boy, opposing other boys in teams entails affiliation within the team. (The most dramatic instance of male affiliation resulting from conflict with others is bonding among soldiers, a phenomenon explored by Norman 1990.) By the same token, girls' efforts to support their friends necessarily entail exclusion of or opposition to other girls. This emerges in Hughes' 1988 study of girls playing the game foursquare. The social injunction to be "nice" and not "mean" was at odds with the object of the game, to eliminate players who are then replaced by awaiting players. The girls resolved the conflict, and formed "incipient teams" composed of friends, by claiming that their motivation in eliminating some players was to enable others (their friends) to enter the game. In their terms, this was "nice-mean." This dynamic is also supported by my analysis of the sixth grade girls' conversation: Most of their talk was devoted to allying themselves with each other in opposition to another girl who was not present. So their cooperation (solidarity) also entails opposition (power).

For boys, power entails solidarity not only by opposition to another team, but by opposition to each other. In the videotapes of friends talking, I found that all the conversations between young boys (and none between young girls) had numerous examples of teasing and mock attack. In examining pre-school conversations transcribed and analyzed by Corsaro and Rizzo 1990, I was amazed to discover that a fight can be a way of initiating rather than precluding friendship. In one episode, a little boy intrudes on two others and an angry fight ensues in which they threaten to punch and shoot poop at each other, and to snap a Slinky in each other's faces. By the end of the episode, however, the three boys are playing together amicably. Picking a fight was the third boy's way of joining the play of the other two.

These examples call into question the correlation of aggression and power on one hand, and cooperation and solidarity on the other. Doubt is also cast by a cross-cultural perspective. For example, many cultures of the world see arguing as a

pleasurable sign of intimacy. Schiffrin 1984 shows that among working class men <u>and</u> <u>women</u> of East European Jewish background, friendly argument is a means of being sociable. Frank 1988 shows a Jewish couple who tend to polarize and take argumentative positions, but they are not fighting; they are staging a kind of public sparring, where both fighters are on the same team. Byrnes 1986 claims that Germans find American students uninformed and uncommitted because they are reluctant to argue politics with new acquaintances. For their part, Americans find German students belligerent because they start arguments about American foreign policy with Americans they have just met.

Greek conversation provides an example of a cultural style that places more positive value, for women and men, on dynamic opposition. Kakava 1989 replicates Schiffrin's findings by showing how a Greek family enjoy opposing each other in dinner table conversation. In another study of modern Greek conversation, Tannen and Kakava 1989 find speakers routinely disagreeing when they actually agree, and using diminutive name forms and other terms of endearment -- markers of closeness -- just when they are opposing each other. In the following excerpt, for example, I express agreement with my interlocutor, an older Greek woman who has just told me that she complained to the police about a construction crew that illegally continued drilling and pounding through the siesta hours:

Deborah: Echete dikio.
Stella: Ego <u>echo</u> dikio. Kopella mou, den xero an echo dikio i den echo dikio. Alla ego yperaspizomai ta symferonta mou kai ta dikaiomata mou.

Deborah: You're right.
Stella: I <u>am</u> right. My dear girl, I don't know if I'm right or I'm not right. But I am watching out for my interests and my rights.

Stella disagrees with my agreement with her by reframing my agreement in her own terms rather than simply accepting it by stopping after "I <u>am</u> right." She also marks her divergence from my frame with the endearment "kopella mou" (literally, "my girl").

In another conversation, presented by Kakava as typical of her family's sociable argument, the younger sister has said that she cannot understand why the attractive young woman who is the prime minister Papandreou's girl friend would have an affair with such an old man. The older sister, Christina, argues that the woman may have felt that in having an affair with the prime minister she was doing something notable. Her sister replied, "Poly megalo timima re Christinaki na pliroseis pantos" ("It's a very high price to pay, Chrissie, anyway"). I am using the English diminutive form "Chrissie" to reflect the Greek diminutive ending -<u>aki</u>, but the particle <u>re</u> cannot really be translated; it is simply a marker of closeness that is typically used when disagreeing, as in the ubiquitously heard "Ochi, re" ("No, <u>re</u>").

Conclusion. The intersection of language and gender provides a rich site for analyzing how power and solidarity are created in discourse. But prior research in this area evidences the danger of linking linguistic forms with interactional intentions such as dominance. In trying to understand how speakers use language, we must consider the context (in every sense, including at least textual, relational, and institutional

constraints), speakers' conversational styles, and, most crucially, the interaction of their styles with each other.

Attempts to understand what goes on between women and men in conversation are muddled by the ambiguity of power and solidarity. The same linguistic means can accomplish either, and every utterance combines elements of both. Scholars, however, like individuals in interaction, are likely to see only one and not the other, like the picture that cannot be seen for what it is -- simultaneously a chalice and two faces -- but can only be seen alternatively as one or the other. In attempting the impossible task of keeping both images in focus at once, we may at least succeed in switching from one to the other fast enough to deepen our understanding of power and solidarity as well as communication between women and men.

Note

1. Following the oral presentation of this paper, both Gary Holland and Michael Chandler pointed out that the prisoner may be heard as implying the second part of the guard's statement: "and you're a pile of shit."

References

Becker, A.L. 1982. Beyond translation: Esthetics and language description. Contemporary perceptions of language: Interdisciplinary dimensions. Georgetown University Round Table on Languages and Linguistics 1982, ed. by Heidi Byrnes, 124-138. Washington, DC: Georgetown University Press.
Becker, A. L., and I Gusti Ngurah Oka. 1974. Person in Kawi: Exploration of an elementary semantic dimension. Oceanic Linguistics 13:229-255.
Bellinger, David, and Jean Berko Gleason. 1982. Sex differences in parental directives to young children. Sex Roles 8:1123-39.
Brown, Roger, and Albert Gilman. 1960. The pronouns of power and solidarity. Style in language, ed. by Thomas Sebeok, 253-276. Cambridge, Mass.: M.I.T. Press.
Brown, Penelope, and Stephen Levinson. [1978]1987. Politeness: Some universals in language usage. Cambridge: Cambridge University Press.
Byrnes, Heidi. 1986. Interactional style in German and American conversations. Text 6:2.189-206.
Coates, Jennifer. 1986. Women, men and language. London: Longman.
Conley, John M., William M. O'Barr, and E. Allen Lind. 1979. The power of language: Presentational style in the courtroom. Duke Law Journal vol. 1978 no. 6.
Corsaro, William, and Thomas Rizzo. 1990. Disputes in the peer culture of American and Italian nursery school children. Conflict talk, ed. by Allen Grimshaw, 21-65. Cambridge: Cambridge University Press.
Dorval, Bruce (ed.) 1990. Conversational coherence and its development. Norwood, NJ: Ablex.
Fasold, Ralph W. 1990. The sociolinguistics of language. Oxford: Basil Blackwell.
Frank, Jane. 1988. Communicating "by pairs": Agreeing and disagreeing among married couples. Ms., Georgetown University.

Friedrich, Paul. 1972. Social context and semantic feature: The Russian pronominal usage. Directions in sociolinguistics, ed. by John J. Gumperz and Dell Hymes, 270-300. New York: Holt, Rinehart, Winston.

Gleason, Jean Berko. 1987. Sex differences in parent-child interaction. Language, gender, and sex in comparative perspective, ed. by Susan U. Philips, Susan Steele, and Christine Tanz, 189-99. Cambridge: Cambridge University Press.

Gumperz, John J. 1982. Discourse strategies. Cambridge: Cambridge University Press.

Hughes, Linda A. 1988. "But that's not really mean": Competing in a cooperative mode. Sex Roles 19:11/12.669-87.

Kakava, Christina. 1989. Argumentative conversation in a Greek family. Paper presented at the Annual Meeting of the Linguistic Society of America, Washington, DC.

Keenan, Elinor. 1974. "Norm-makers, norm-breakers: Uses of speech by men and women in a Malagasy community." Explorations in the ethnography of speaking, ed. by Richard Bauman and Joel Sherzer, 125-43. Cambridge: Cambridge University Press.

Komarovsky, Mirra. 1962. Blue-collar marriage. New York: Vintage.

Lakoff, Robin. 1975. Language and woman's place. New York: Harper and Row.

Leet-Pellegrini, H. M. 1980. Conversational dominance as a function of gender and expertise. Language: Social psychological perspectives, ed. by Howard Giles, W. Peter Robinson, and Philip M. Smith, 97-104. Oxford: Pergamon.

Maltz, Daniel N., and Ruth A. Borker. 1982. A cultural approach to male-female miscommunication. Language and social identity, ed. by John J. Gumperz, 196-216. Cambridge: Cambridge University Press.

Moore, Lorrie. 1989. You're ugly, too. The New Yorker, July 3, 1989, 30-40.

Norman, Michael. 1990. These good men: Friendships forged from war. New York: Crown.

Ong, Walter J. 1981. Fighting for life: Contest, sexuality, and consciousness. Ithaca: Cornell University Press; Amherst: University of Massachusetts Press.

Pinter, Harold. 1988. Mountain language. New York: Grove Press.

Sattel, Jack W. 1983. Men, inexpressiveness, and power. Language, gender and society, ed. by Barrie Thorne, Cheris Kramarae, and Nancy Henley, 119-24. Rowley, MA: Newbury House.

Schiffrin, Deborah. 1984. Jewish argument as sociability. Language in Society 13:3.311-335.

Scollon, Ron. 1982. The rhythmic integration of ordinary talk. Analyzing discourse: Text and talk. Georgetown University Round Table on Languages and Linguistics 1981, ed. by Deborah Tannen, 335-349. Washington, DC: Georgetown University Press.

Scollon, Ron. 1985. The machine stops: Silence in the metaphor of malfunction. Perspectives on silence, ed. by Deborah Tannen and Muriel Saville-Troike, 21-30. Norwood, NJ: Ablex.

Scollon, Ron, and Suzanne B.K. Scollon. 1981. Narrative, literacy and face in interethnic communication. Norwood, NJ: Ablex.

Shuy, Roger W. 1982. Topic as the unit of analysis in a criminal law case. Analyzing discourse: Text and talk. Georgetown University Round Table on Languages and Linguistics 1981, ed. by Deborah Tannen, 113-26. Washington, DC: Georgetown University Press.

Spender, Dale. 1980. Man made language. London: Routledge and Kegan Paul.

Tannen, Deborah. 1981. Indirectness in discourse: Ethnicity as conversational style. Discourse Processes 4(1981):3.221-238.

Tannen, Deborah. 1984. Conversational style: Analyzing talk among friends. Norwood, NJ: Ablex.

Tannen, Deborah. 1985. Silence: Anything but. Perspectives on silence, ed. by Deborah Tannen and Muriel Saville-Troike, 93-111. Norwood, NJ: Ablex.

Tannen, Deborah. 1986. That's not what I meant!: How conversational style makes or breaks your relations with others. New York: William Morrow, Ballantine.

Tannen, Deborah. 1987. Repetition in conversation: Toward a poetics of talk. Language 63:3.574-605.

Tannen, Deborah. 1989a. Talking voices: Repetition, dialogue and imagery in conversational discourse. Cambridge: Cambridge University Press.

Tannen, Deborah. 1989b. Interpreting interruption in conversation. Papers from the 25th Annual Regional Meeting of the Chicago Linguistic Society, ed. by Bradley Music, Randolph Graczyk, and Caroline Wiltshire, 266-87. Chicago: Department of Linguistics, University of Chicago.

Tannen, Deborah. 1990a. Gender differences in conversational coherence: Physical alignment and topical cohesion. Conversational coherence and its development, ed. by Bruce Dorval, 167-206. Norwood, NJ: Ablex.

Tannen, Deborah. 1990b. Gender Differences in Topical Coherence: Creating Involvement in Best Friends' Talk. Discourse Processes 13:1.73-90.

Tannen, Deborah. 1990c. You just don't understand: Women and men in conversation. New York: William Morrow.

Tannen, Deborah, and Christina Kakava. 1989. Power and solidarity in men's and women's conversations in modern Greek. Paper presented at a conference entitled "Power/Freedom: Politics, Social Life, and the Arts in Modern Greece." Minneapolis, Minnesota, October 19-22, 1989.

West, Candace, and Don H. Zimmerman. 1983. Small insults: A study of interruptions in cross-sex conversations between unacquainted persons. Language, gender and society, ed. by Barrie Thorne, Cheris Kramarae, and Nancy Henley, 103-117. Rowley, MA: Newbury House.

Good and Bad News in Formalizing Generalized Implicatures

Jacques Wainer   Anthony Maida
Pennsylvania State University

This paper examines the pros and cons of a model-theoretic formalization of a class of generalized implicatures. We concentrate on generalized implicatures that derive from the assumption that the speaker is observing the maxim of Quantity. There is a family of logics, called nonmonotonic logics, which, at a theoretical level, seem well suited to modeling generalized implicatures. However, when one develops the details, the approach loses its elegance. Difficulties arise because current nonmonotonic logics are not fine-grained enough in the kinds of differential implicatures that they can model. In particular, given two distinct logical forms which are logically equivalent, nonmonotonic logics are forced to derive the same conclusion from each logical form.

Gazdar's (1979) formalization of generalized implicatures has two differences from that examined in this paper. First, Gazdar's approach is centered on the lexical items, whereas our approach is based on model-theoretic concepts. The second difference is that we plan to concentrate on a form of generalized implicature that arises from the assumption that the speaker mentioned all relevant entities in the utterance. This form of implicature does not seem to be easily incorporated into Gazdar's framework.

The next section states why nonmonotonic logic should be an appropriate framework with which to model generalized implicatures. Section 2 discusses a particular form of generalized implicature, the only-implicatures, and we present the similarities between this implicature and circumscription, a particular form of nonmonotonic logic. We then apply the approach to some simple examples with some relative success. Section 3 discusses the major limitations of the approach. We discuss a way of formalizing nonmonotonic logic through semantic concepts which allow us to prove that distinct but logically equivalent logical forms lead to identical conclusions. We then come up with some simple examples for which, we claim, there is no simple way of coming up with distinct non-logically equivalent logical forms. Finally we discuss some of the possible directions for future research in the area of formalizing these kinds of implicatures.

1   The Approach

Our approach is to embed the assumptions stated in Grice's maxim of Quantity into a logic $\mathcal{L}$. We then specify the generalized implicatures of an utterance in terms of

logical entailment within $\mathcal{L}$. In other words, the implicatures of an utterance are calculated by performing some logical deduction from the semantic content of the sentence-utterance.

The logic $\mathcal{L}$, if it is to model implicatures, must be nonmonotonic. Monotonicity is the property of a logic that if some formula is entailed from a set of assumptions, then the same formula will also be entailed for any other set of assumptions that contains the original set. If we use the symbol $\models$ to denote logical entailment, Γ to denote the original set of assumptions, and Δ for the additional assumptions, then expression (1) will hold for any monotonic logic.

(1)   if $\Gamma \models \alpha$ then $\Gamma \cup \Delta \models \alpha$

Traditional first-order logics, modal logics, intentional logics, and so on, are all monotonic. On the other hand, for a nonmonotonic logic it is possible that

(2) $\Gamma \models \alpha$ but $\Gamma \cup \Delta \not\models \alpha$

Nonmonotonicity is necessary to account for the phenomenon of cancelability of implicatures. For example, the utterance (3.a) carries (3.b) as an implicature. But (3.c) does not carry the same implicature.

(3)   a. Some of the girls are sick.
       b. Some of the girls are sick and not all of the girls are sick.
       c. Some, and maybe all of the girls are sick.

If we assume that the function [ ] takes as argument a language expression and denotes the expression's semantic content in the language of $\mathcal{L}$, and if the logic $\mathcal{L}$ correctly models implicatures, then the semantic content of (3.b) should be derivable in $\mathcal{L}$ from the semantic content of (3.a). In a more mathematical notation:

(4)   { [some of the girls are sick] } $\models_{\mathcal{L}}$ [not all of the girls are sick]

But if one assumes compositionality, then one must acknowledge the semantic content of (3.c) is

(5)   [some of the girls are sick] $\wedge$ [maybe all girls are sick]

Thus, nonmonotonicity in a logic is necessary if that logic is to account for the cancelability of implicatures. Furthermore, this framework allows us to model some forms of cancelability derived from context information, as opposed to the form of cancelability by adding extra clauses shown above. For example

(6)   The eggs are in the garden or in the living room.

generally implicates that as far as the speaker knows the eggs could be in either place and the speaker does not know which. But given the information that the speaker is the sponsor of an Easter egg hunt, and that he knows where the eggs are, the implicature that he does not know where the eggs are is canceled. Common knowledge is easily represented as formulas in the same logical language one uses to represent the semantic content of sentences. Thus the derivation from the logical

form of (6) of some formula that states that the speaker does not know where the eggs are, is canceled with the addition of the assumption that the speaker knows where the eggs are. Other forms of context information, like focus, may also contribute to cancel implicatures. But focus is unlikely to be easily modeled by this approach. This is because it is unclear how to represent focus as a set of formulas in a logical language.

This is the proper place to discuss our assumptions. First we assume that there is such a thing as the literal meaning of utterances. Of this literal meaning, we are interested in the components that correspond to truth functional concepts, which we called above as the propositional content of the utterance. But the most important assumption is that the propositional content, and indeed the literal meaning of an utterance cannot be canceled if the utterance is being used literally. In other words, by adding more clauses to a literal utterance one cannot cancel any part of the original utterance's propositional content. This is the key assumption in the discussion of what is the propositional content of some particular sentences later on.

We will represent the propositional content of an utterance by a formula in some standard logical language, in particular, for most of the examples in this paper we use a first-order language to represent the propositional content of utterances, except for examples where one must be explicit about the state of knowledge of the speaker and then we use a quantified modal logic language.

2 Good News

We can show that the approach suggested in the previous section is indeed able to generate some of the generalized implicatures of utterances. In particular, we concentrate on only-implicatures.

Only-implicatures

We claim that there is a class of generalized implicatures, which we call ONLY-IMPLICATURES, which derives from a consequence of the Quantity maxim, namely that the speaker mentioned all relevant entities in the utterance. Typical examples of only-implicatures are sentences whose implicatures are obtained by adding the particle *only* in some selected places. Let us see some unproblematic examples, where the first sentence is the utterance and the second is what is implicated:

(7)   a. John has two children.
        b. John has ONLY two children.

(8)   a. The flag is white and blue.
        b. The flag is ONLY white and blue.

(9)   a. Tom and Jerry moved the sofa.
        b. ONLY Tom and Jerry moved the sofa.

(10)   a. Mary fell from the ferry boat.
        b. ONLY Mary fell from the ferry boat.

So for example, in (7.a) the speaker mentions the existence of two of John's children. If one assumes that the speaker mentioned all of the relevant entities, and relevance here seems to mean the entities that are children of John's, then one can conclude the total number of John's children. Actually, all statements above must be qualified by the speaker's knowledge. One should read the explanation above as: since the speaker mentioned the existence of two of John's children, one can assume that that is the total number of John's children as far as the speaker knows. In the remainder of this paper, we will leave out the references to the speaker's knowledge, and the proper qualification must be assumed by the reader. Only when the knowledge of the speaker (or lack of it) becomes important as part of the literal meaning of a sentence, or as part of the implicature, will we explicitly refer to it.

Different aspects of only-implicatures have been discussed in the literature before. For example, the implicature described in (7) above can be explained by the scalar implicature formalizations of Horn (1972) and of Gazdar (1979). Also, only-implicatures seem to be related with Harnish (1976) all-implicatures from where we take example (8) above. But the scope of only-implicatures extends beyond those approaches, as examples (9) and (10) show. In particular, it seems unlikely that Gazdar's formalization, which is centered on lexical items, could be extended to include examples (9) and (10) because there is no lexical item that can be said to be the trigger of the implicature.

However, the existence of only-implicatures as a general phenomenon is debatable (Fauconnier 1990). The idea that all relevant entities have been mentioned would allow for unwanted implicatures, for example, from (7.a) one should conclude that

(11)   ONLY John has only two children.

which is clearly not a generalized implicature of (7.a). The extended version of this paper (Wainer and Maida 1990) addresses this point showing that only-implicatures can still be considered as an instance of generalized implicatures that are further canceled by stereotypical knowledge about the number of children one usually has.

The assumption that all relevant objects have been mentioned is also present in the field of formalizing common sense reasoning. In the attempt to provide a formalization to the common sense inference described in (12), McCarthy(1980) created a family of nonmonotonic logics, called CIRCUMSCRIPTION, that capture the assumption that all objects have been mentioned.

(12)   Most birds fly.
        Penguins do not fly and they are birds.
        Tweety is a bird.
        _____
        Tweety flies

McCarthy's solution is to make explicit a predicate of being abnormal-in-the-aspect-of-flying, which all penguins satisfy. In other words, all birds that are not

abnormal-in-the-aspect-of-flying do fly and all penguins are abnormal-in-the-aspect-of-flying. Given that Tweety is a bird, and if we assume that given what is known, abnormal-in-the-aspect-of-flying satisfies the smallest number of birds, then one may conclude that Tweety can fly. More simply, if Tweety is not known to be abnormal-in-the-aspect-of-flying, assume it is not.

Before we discuss the details of circumscription, we would like to point out the connections (at this moment only intuitive) between the problem stated in (12) and Grice's quantity maxim. It seems to be exactly because one assumes that statement 'Tweety is a bird' was as informative as required for the purposes of the problem that one would want to conclude that Tweety flies.

Circumscription

Given a set Γ of assumptions from which one would like to derive conclusions, predicate circumscription is accomplished by adding a second order axiom (called the CIRCUMSCRIPTION SCHEMA) to the set of assumptions 'before' performing the logic derivations. Derivations are then performed in (monotonic) second-order logic. Nonmonotonicity arises from the fact that the schema is added 'after' all assumptions have been collected, and so if the set of assumptions is changed to $\Gamma \wedge \beta$, the circumscription schema itself would be different, and this different axiom could invalidate conclusions that have been derived using the previous schema.

If P is the predicate one wishes to circumscribe, and $\Gamma(P)$ is the set of assumptions, then the circumscription schema will be

(13) $\forall \phi \ (\Gamma(\phi) \rightarrow P \leq \phi)$

where ϕ is a predicate variable of the same arity of P, $\Gamma(\phi)$ is the same set of assumption where the symbol P has been substituted everywhere by ϕ, and $P \leq \phi$ is an abbreviation for

(14) $\forall x \ (P(x) \rightarrow \phi(x))$ where x is a tuple of variables of the appropriate size.

What (13) states is that if the set of conclusions holds for another predicate substituted in the place of P then the new predicate is at least as large as P. Or in other words, it states that P has the smallest extension compatible with Γ.

More complex forms of circumscription have been developed: POINTWISE CIRCUMSCRIPTION (Lifschitz 1986); PARALLEL CIRCUMSCRIPTION and PRIORITIZED CIRCUMSCRIPTION (McCarthy 1980). These formalisms propose different forms in which the minimization of the extension of a predicate is carried out.

Examples

As it was shown above, only-implicatures seem to have strong connections with circumscription. We will now show that circumscription will generate correct only-implicatures for some cases. However, to calculate the implicatures by applying circumscription, one must first have in hand the propositional content of the sentence. It is to the discussion of the propositional content of (15) that we now turn.

(15)   John has two children.

As we stated earlier, we will represent the propositional content of sentences by formulas first order logic. We claim that the propositional content of (15) is fully captured by the formula

(16)   ∃xy [child(x,John) ∧ child(y,John) ∧ x ≠ y]

which states that there are two different entities, each of which is a child of John. What should be made clear is that the formula does not specify the exact number of John's children, but it sets a lower bound to the number of John's children as two. Or in other words, the propositional content of (15) is a formula that one would associate with the import of 'John has at least two children.' To defend this claim we must show that there is no stronger formula that can capture the propositional content of (16).

One could reasonably hypothesize that the propositional content of (15) is that John has exactly two children, represented by expression (17) below.

(17)   ∃xy [child(x,John) ∧ child(y,John) ∧ x ≠ y
              ∧ ∀z [child(z,John) → (x = z ∨ y = z)]]

The reason why (17) seems too strong to be the propositional content of (15) derives from a compositionality argument, coupled with the assumption that literal meaning is not cancellable. In particular, the literal meaning of the first clause of expression (18) below

(18)   John has two children, in fact he has three.

should be identical to the literal meaning of (15). The propositional content of the second clause of (18) above should be analogous to (17) except that the expression should capture the fact that there are three, instead of two children. If we assume that the truth functional import of *in fact* is the same as the logical *and* then the propositional content of (18) should be the same as that of expression (19), but (19) is inconsistent and (18) is not.

(19)   John has exactly two children, in fact he has exactly three.

The claim that *in fact* has the same truth functional import as *and* is debatable (Wilensky 1990). The expression *in fact* in some way asserts both clauses it links, although it has some other pragmatic imports. But since literal meaning cannot be canceled, none of these extra factors will cancel the fact that both clauses are being asserted, or in other words, *in fact* has at least the same truth functional import as *and*. Actually, the main source of limitations for this assumption is that *in fact* is being modeled by a symmetric operator when it is really not a symmetric connector. The consequence is that the theory will make incorrect predictions for anomalous expressions like

(20)   *John has three children, in fact he has two.

We have shown that the propositional content of (15) cannot be (17), and that it must be (16). This result is not as counter intuitive as it may seem in a first analysis. The fact that (17) is not the propositional content of (15) states that the fact that John has only two children is really an implicature and not an entailment of (15).

Given that we accept that (16) is the propositional content of (15), we must now specify a logic such that (15) would entail the proper only-implicature. This is accomplished by circumscribing the predicate $\lambda(x)child(x, John)$ (being a child of John's) in (16). The formula

(21)   ∃xy [child(x,John) ∧ child(y,John) ∧ x ≠ y
                  ∧ ∀z [child(z,John) → (x = z ∨ y = z)]]

which roughly states that John has only two children, will be derivable from the circumscription schema.

To be more precise, what is done is the circumscription of a dummy predicate symbol Q defined to be equivalent to $\lambda(x)child(x, John)$. In this case the set of assumptions is

(22)   ∀z [Q(z) ↔ child(z,John)] ∧
                  ∃xy [child(x,John) ∧ child(y,John) ∧ x ≠ y]

and Q is circumscribed in (22) allowing the extension of child to vary as needed. From now on, when we mention the circumscription of a complex predicate, one that is constructed using the λ-operator, we mean a circumscription carried along the lines described above.

Implicature derived from circumscription are nonmonotonic, as the example below will show. The propositional content of

(23)   John has two children, in fact he has three.

will be equivalent to

(24)   ∃xy [child(x,John) ∧ child(y,John) ∧ x ≠ y] ∧
                  ∃xyz [child(x,John) ∧ child(y,John) ∧ child(z,John)
                      ∧ x ≠ y ∧ x ≠ z ∧ y ≠ z]

Again, circumscribing $\lambda(x)child(x, John)$ in (24) above would result in a formula that states John has only three children, which is the correct implicature.

A similar procedure when applied to examples below

(25)   a. The flag is white and blue.
              b. John moved the sofa and the bed.

would generate the correct implicatures in (26)

(26)   a. The flag is only white and blue.
              b. John moved only the sofa and the bed.

In the first example what is circumscribed is $\lambda(x)color(x,y)$ (being the color of something). In the second example the predicate $\lambda(x)move(John,x)$ (being moved by John) is circumscribed.

On the other hand, the limitations of those results are clear. The ad hoc way in which the predicate to be circumscribed has to be picked must be explained and further research is being done in the direction of automatizing this choice. And the results are very sensitive to that choice. For example, going back to (15) and its logical form (16), if instead of circumscribing $\lambda(x)child(x,John)$ we had circumscribed the predicate symbol child the resulting conclusion would have been a formula that has the same import of (27) below.

(27)   There is only one parent, which is John, and he has exactly two children.

3   Bad News

Although the examples developed in the previous section suggest that the approach has some potential, we have found examples where the idea of modeling implicatures as entailment within a nonmonotonic logic cannot succeed, for any logic of interest. The reason for the failure is that circumscription (actually a larger class of nonmonotonic logics) is insensitive to logical equivalence. In other words, two different formulas that are logically equivalent will entail the same set of conclusions no matter what predicate is being circumscribed or what version of circumscription is being used.

To show that we must first explain this larger class on nonmonotonic formalisms, which contains all forms of circumscription, and then prove that any logic in this class will be insensitive to logical equivalence.

All forms of circumscription can be classified as semantically based logics. This is to say that besides the axiomatic definition one can also give a model-theoretic characterization of the various forms of circumscription. Those model-theoretic characterizations are all based in an idea, first developed by Shoham (1987), of defining an ordering relation among the class of models for a set of formulas. This ordering relation, called PREFERENCE, defines a partial order of models, and allows one to define MAXIMALLY PREFERRED MODELS as models for which there is no other model that is 'more preferred'.

LOGICAL ENTAILMENT is then defined as: a set of formulas Γ logically entails β if and only if all maximally preferred models for Γ also satisfy β. The traditional definition of logical entailment would require that all models that satisfy Γ must also satisfy β. Different logics are created for different definitions of the preference relation $\mathfrak{F}$. For example, the preference relation for the predicate circumscription of a predicate symbol P, would state that M_1 is preferred to M_2 if they agree with each other in the denotation of every constant symbol and every predicate symbol, except that the extension of P in M_1 is smaller than in M_2.

Given the above description of the semantically-based nonmonotonic logics, which includes all forms of circumscription, we can state the theorem that logically

equivalent formulas would yield the same conclusion under any of those logics.

The proof is based on the fact that logically equivalent formulas are satisfied by the same models. Therefore, under any preference criteria, the set of maximally preferred models are the same for both formulas. Since the logical entailments of a formula are exactly the formulas that are satisfied by all maximally preferred models, and logically equivalent formulas have the same set of maximally preferred models, then both formulas will have the same set of entailments.

Given these facts we can now discuss some examples where the sentences would have equivalent logical forms but have different implicatures. Let us compare the sentences (28) and their implicatures in (29)

(28)   a. John has two children.
       b. John has at least two children.

(29)   a. John has only two children.
       b. John has two children and possibly more.

We have argued that the propositional content of (28.a) is captured by expression (16), which also seems to be the propositional content of (28.b). In fact, (16) captures exactly what one would expect to the the content of (28.b): a statement on the lower bound of the number of John's children. But if this is so, then both sentences in (28) have logically equivalent literal meanings and there is no nonmonotonic logic that can derive the correct implicature for both sentences simultaneously.

It is possible that we have given up on expression (28.b) too easily and that its logical form does contain more information than that shown in (16). We have not accounted for the contribution of the phrase *at least* in the literal meaning of the sentence. Perhaps, it means that the speaker considers possible that John has more than two children, and this should be part of the literal meaning of the sentence. In other words, the propositional content of (28.b) would be captured by the import of

(30)   The speaker knows that there are (at least) two things that
          are John's children and it is possible considering his knowledge
          that there is (at least) a third child.

The extended version of this paper (Wainer and Maida 1990) shows that if (30) is represented by a modal logic formula (Hintikka 1962), then

(31)   John has at least two children, in fact he has exactly two.

would be contradictory.

Thus, under reasonable theories of propositional content the pair of sentences in (28) would yield logically equivalent logical forms. Thus it is impossible to construct a nonmonotonic logic that simultaneously derive the correct implicature for each sentence..

Other sentences which under a reasonable theory of propositional content would have semantic interpretations which are logically equivalent but whose correspondent sentence-utterances carries different implicatures is the pair below:

(32)   a. John has two or three children.
        b. John has two or more children.

In a nutshell, the problem with the sentences above is that *three* is translated to 'there are (at least) three things that satisfies ...' and *more*, which we take to be an ellipsis of 'more than two', would have an equivalent translation.

4   Conclusions

We presented an approach to modeling generalized quantity implicatures as entailment in some nonmonotonic logic. This approach seems natural and has potential. We have shown that circumscription does derive some of the correct only-implicatures. Some examples of clausal implicatures, in particular the interaction between modal operators like *knowledge* and logical connectives, although not discussed in this paper are also being investigated by the authors.

On the other hand, we have identified a major limitation of the approach: its inability to distinguish between logically equivalent formulas. We have pointed out examples where a reasonable interpretation of the semantic content of some sentences generates logically equivalent logical forms for sentences that carry different implicatures.

Future research points in the direction of incorporating the calculation of generalized implicatures in the semantic interpretation itself along the lines of Karttunen and Peters (1979) work in conventional implicatures. There is evidence that the derivation of generalized implicatures is centered more on lexical and syntactical items than on semantic concepts. For example, the presence of the expression *at least* in (28.b) is responsible for a different set of implicatures in comparison with (28.a). Other evidence comes from the examples below:

(33)   a. John sees two children.
        b. John seeks two unicorns.

The implicature for both cases is very similar:

(34)   a. John sees only two children.
        b. John seeks only two unicorns.

but the logical forms for sentences (33) are much different. It is more reasonable to assume that the implicature derivation process is in the semantic interpretation function. This is because the presence of the word *two*, independent on whether it appears within an intentional context or not, generates the implicature. This theory seems more plausible than to assume that some logic, working from very different logical forms, would derive such similar implicatures.

References

Fauconnier, Gilles. 1990. Personal communication, 16th Annual Meeting of the BLS.

Gazdar, Gerald. 1979. *Pragmatics: Implicature, Presupposition and Logical Form.* Academic Press, New York.

Grice, Paul. 1975. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics 3: Speech Acts*, Academic Press, New York.

Harnish, Robert M. 1976. Logical form and implicature. In T. Bever, J. Katz, and T. Langendoen, editors, *An Integrated Theory of Linguistic Ability*, pages 313–392, Thomas Y. Crowell Company, New York.

Hintikka, Jaakko. 1962. *Knowledge and Belief.* Cornell University Press.

Horn, L. 1972. On the semantic properties of the logical operators in English. Unpublished doctoral dissertation, UCLA, Los Angeles.

Karttunen, Lauri and Peters, Stanley. 1979. Conventional implicature. In Choon-Kyu Oh and David A. Dinneen, editors, *Syntax and Semantics: Presupposition*, Academic Press, New York.

Lifschitz, V. 1986. Pointwise circumscription: preliminary report. In *Proc. American Assoc. for Artificial Intelligence*, pages 406–410, Philadelphia, PA.

McCarthy, John. 1980. Circumscription—a form of nonmonotonic reasoning. *Artificial Intelligence*, (13):27–39.

Shoham, Y. 1987. A semantical approach to nonmonotonic logics. In *Proc. of the Tenth International Joint Conference on Artificial Intelligence*, pages 388–392.

Wainer, Jacques and Maida, Anthony. 1990. Good and Bad News in Formalizing Generalized Implicatures (extended version). Technical Report CS–90–11 , The Pennsylvania State University, Department of Computer Science, 333 Whitmore Lab, University Park, PA 16802.

Wilensky, Robert. 1990. Personal communication, 16th Annual Meeting of the BLS.

PAPER PRESENTED AT

BLS 15

PARASESSION

ON

THEORETICAL ISSUES

IN

LANGUAGE

RECONSTRUCTION

Bulging Monosyllables: Areal Tendencies in Southeast Asian Diachrony

James A. Matisoff
University of California, Berkeley

1.   The complex monosyllable and intersegmental slopover.[1]

Most language families of East and SE Asia are "monosyllabic" (in the sense that their morpheme stocks, not necessarily their words) consist overwhelmingly of monosyllables.[2] As I observed long ago, "There is something about the tightly structured nature of the syllable in monosyllabic languages which favors the shift in contrastive function from one phonological feature of the syllable to another."[3]   There is indeed much to be shifted around in these languages. We are not talking about such puny little (C)V(C) syllables as in Indonesian **rambut** 'hair' or Japanese **kokoro** 'heart',[4] but rather such robust entities as Vietnamese **nguyên** [ŋwyən] 'spring; source', Hmong **npluag** 'chaff' [the -g is a tonemark indicating a falling pitch with breathy voice], Thai **khrîaŋ** 'gear; apparatus', or Written Tibetan **bsnyigs** 'sediment'.

Sometimes there is a morphological basis for a syllable's phonological complexity, as in the Aslian branch of Mon-Khmer (Malaya), where various derivational processes involve the infixation of a copy ("incopyfixation") of the syllable-final consonant, leading sometimes to spectacular consonant clusters:[5]

(1) *with incopyfix of final alone (roots with initial clusters)*
          Ci Cii V Cf   ---->   Ci Cf Cii V Cf
    Cheq Wong (N. Asl.) **hwɛc** 'whistle' / **hcwɛc** 'whistling'

(2) *with root-external infix plus incopyfix*
  (a) simple initial
          Ci V Cf   ---->   Ci Cf N V Cf
    Semai (C. Asl.) **te:w** 'river' (mass) / **twne:w** [tuni:w] 'river' (count)
  (b) cluster initial
          Ci Cii V Cf   ---->   Ci N Cf Cii V Cf
    Semai **sla:y** 'swidden' (mass) / **snyla:y** [snila:y] 'id. ' (count)

(3) *with root-external prefix plus incopyfix*
          Ci V Cf   ---->   N Cf Ci V Cf
    Semelai (S.Asl.) **kap** 'bite' / **npkap** 'biting'

(4) *with reduplication of the initial and incopyfixation of the final*

	Ci V Cf		---->	Ci Cf Ci V Cf

Bateg (N. Asl.)   **kwᴄ** 'grate' / **kᴄkwᴄ** 'is grating'

Jah Hut (C. Asl.)   **ᴏaʔ** 'eat' / **ᴏʔᴏaʔ** 'is eating'

Semai   **laal** 'stick out tongue' / **lllaal** [ləlla:l] 'id.' *(prog.)*

Temiar (C. Asl.)   **hᴐᴐh** 'follow' / **hhhᴐᴐh** [hɛhhᴐᴐh] 'is following'
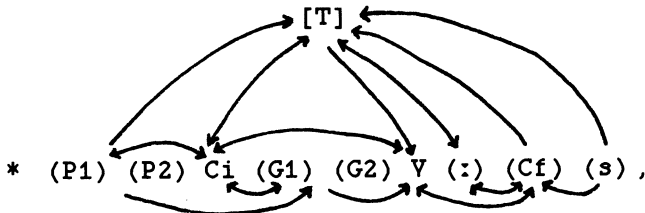
Note that these languages do not shrink from applying this process even when the root-initial and -final consonants are identical (as in the last two examples), even though this leads to four occurrences of the same consonant in the syllable!

Tibeto-Burman has never had a system of infixes[6], yet the Proto-Tibeto-Burman [PTB] syllable canon was certainly complex enough:

$$[T]$$

$$* \ (P1) \ (P2) \ Ci \ (G1) \ (G2) \ V \ (:) \ (Cf) \ (s) \ ,$$

where P = prefix, C_i = initial consonant, G = glide, V = vowel, : = vowel length, C_f = final consonant, s = suffixal -s, and T = tone (which I consider to have been phonetically present but phonologically redundant at the proto-stage).[7] Both synchronic and diachronic evidence shows that virtually all parts of such syllables are capable of influencing every other. We may crudely indicate this by inserting ligatures in the above formula:

$$* \ (P1) \ (P2) \ Ci \ (G1) \ (G2) \ V \ (:) \ (Cf) \ (s) \ ,$$

A few examples of some of these types of interrelationship:[8]

(a) $C_f > V$

PLB	Lahu		PLB	Lahu
*-am	→ -o		*-ap	→ -oʔ
*-an	→ -e		*-at	→ -eʔ
*-aŋ	→ -ᴐ		*-ak	→ -aʔ

```
e.g. fathom  *lam¹     lò        rub    *sap       š$\hat{\text{o}}$?
     slave   *gywan¹   cè        leech  *k-r-wat   vè?
     you     *naŋ¹     nɔ̀       rat    *k-r-wak   fâ?
```

Particularly interesting is the "circular" development shown by the PTB rhyme[9] *-ik in the history of Burmese, where the feature of palatality has shifted back and forth from the vowel to the final consonant to the vowel:

PTB *-ik ⇁ W[ritten] B[urmese] -ac ⇁ Mod. Bs. I?,
e.g. PTB *tsik 'joint' (cf. Written Tibetan [WT] tshigs) ⇁ WB chac, Mod. Bs. hsI?

(b) G > Ci

```
*k  ⇁ Lh. qh    PLB *ka²   ⇁ Lh. qhâ  'bitter'
*kr ⇁ Lh. kh    PLB *krəw¹ ⇁ Lh. khɔ  'horn'
*ky ⇁ Lh. ch    PLB *kyəw¹ ⇁ Lh. chɔ  'sweet'
*kw ⇁ Lh. ph    PLB *kwəy² ⇁ Lh. phɛ́ 'dog'
```

(c) G > V

PLB	Lahu	Example
*-a	-a	*n-da¹ 'fern' ⇁ Lh. dà
*-ya	-ɛ	*bya² 'bee' ⇁ Lh. pɛ̂
*-wa	-u	*twa¹ 'handspan' ⇁ Lh. thu

(d) Ci > G

The initial consonant may provoke an assimilatory change in a following glide, as in the following example of "rhinokinesis":

PTB *s-myak 'eye' ⇁ Gyarong təɲak
     [+nas] [-nas]   >   [+nas] [+nas]
            [+pal]   [+pal]

(e) Ci <> V

The interinfluence of initial consonant and vowel is well illustrated, e.g. on the allophonic level in Lahu. Least remarkable are unidirectional cases like Lh. /n/ --⇁ [ñ] /---i, e.g. /nī/ [ñī] 'penis', where the high front vowel palatalizes the preceding consonant -- similar rules might be expected in any language family. More striking are *bidirectional* instances of mutual conditioning of the initial and the vowel, such as obtains with four labials before /u/ and five palatals before /ɨ/:

```
/p ph b m/ ⇁ [pf pfh bv mv] / -------u
/c ch j š y/ ⇁ [ts tsh dz s z] / -------ɨ .
```

These vowels themselves have special realizations here, with /u/ unrounding to [ɯ] and /i̵/ being raised to the "superhigh apical buzz" [ɿ]. Thus, /pú/ 'insect' [pfɯ̂], /yi̵ʔ/ 'sleep' [zɿʔ]. The marriage between initial and vowel is carried to extremes in the syllable /mu/, where the vowel, after affricating the initial, is usually swallowed up entirely, yielding a syllabic nasalized labiodental spirant: /m̥ù/ 'mushroom' [m̥v̩̀].

(f) Ci ... Cf > V

Sometimes both the pre- and post-vocalic consonants contribute phonetic material to the intervening vowel, as in syllables showing "rhinoglottal transfer" in the S. Loloish language Mpi. There are no fewer than nine Mpi syllables which derive from PLB etyma with a nasal root-initial and the rhyme *-ak.[10] In all these cases, the vowel has acquired a complex quality containing both a nasal and a laryngeal component, donated by the pre- and post-vocalic consonants respectively.

> PLB *s-mak 'dream' > Mpi mã̀ʔ   (written maŋʔ)
> PLB *s-nak 'black' > Mpi nã̀ʔ
> PLB *ʔ-ŋak 'open'   > Mpi ŋã̀ʔ

(g) ː > Cf

Vowel length sometimes exerts a crucial influence on a post-vocalic consonant. Proto-Aslian nasals develop into Sabum homorganic stops after short vowels, but into a characteristic type of "decomposed nasals" after long vowels:

> PAslian *kam   ⟶ Sabum kap
>   *kaam   ⟶   kaabm (with "nasal
>   decomposition")

Here the timing of the velic opening is slowed down by the extra mora of the vowel, so that a stop articulation is audible before the final nasal segment.

Similar to this is a phenomenon in Maru (Burmish branch of Lolo-Burmese), where the two open PTB rhymes with long high vowels have developed secondary final stops:[11]

> PTB *-iy or *-əy > Maru -it
> PTB *-uw or *-əw > Maru -uk,

e.g. 'four' PTB *b-liy > Maru bit, 'steal' PTB *r-kuw > Maru khúk

(h)   P > T ; C_l > T ; C_f > T

Perhaps the most remarkable kind of intersegmental feature shuffling is involved in the development of phonemic tone, or *tonogenesis*.[12] In a word, contrastive tone arises as a compensatory mechanism for mergers or losses in the systems of pre- and/or post-vocalic consonants of syllables, especially the loss of a prefix or the neutralization of a voicing contrast in pre-vocalic position, or the loss of a laryngeal segment (-h or -?) after the vowel. The more thoroughly monosyllabic a language family is, the more "tone-prone" it seems to be.
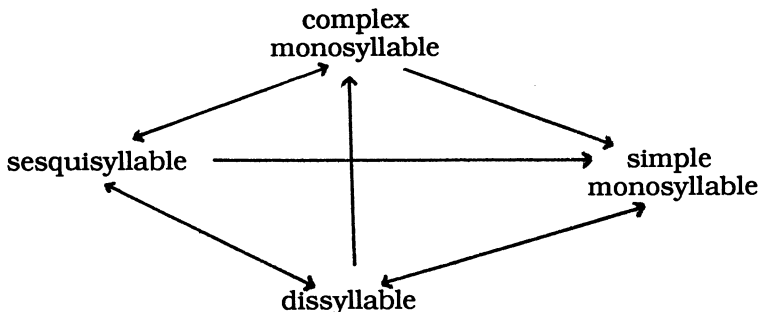
As one example, the high-rising tone /´/ of Lahu can be shown to have arisen through a dissimilatory upheaval in syllables that both began and ended with a "glottal incident", e.g.

> PTB *s-kok 'outer covering' > PLB *?guk > pre-Lahu *?gu? > Lh. qú
> PTB *s-tsik 'joint' > PLB *?dzik > pre-Lahu *?dzi? > Lh. cí

Several branches of Mon-Khmer have developed a similar suprasegmental compensation known as *register* for the loss of a consonantal voicing contrast: a multiplication of the "phonation types" or kinds of voicing with which the vowel may be pronounced. The simplest register systems are two-way, e.g. contrasts between "plain" and "breathy" voice, or between "plain" and "creaky", or "breathy" vs. "creaky". Other languages have 3- or even 4-way systems which rival true tone systems in complexity.[13]

2   The expansion/contraction cycle: monosyllables, dissyllables, and sesquisyllables.

It is my central thesis that the lexemes of Southeast Asian languages undergo a millennial dance from one type of syllable-structure to another, oscillating cyclically among (1) (consonantally) *complex monosyllables* (2)(consonantally) *simple monosyllables* (3) *dissyllabic compounds* or tight collocations and (4) *sesquisyllables*.[14] These developments may be diagrammed crudely as follows:

complex
monosyllable

sesquisyllable   simple
monosyllable

dissyllable

It would be too much to expect that we could trace a given morpheme through this whole cycle -- the period of oscillation is, after all, rather slow. Yet in favorable circumstances the same etymon may indeed be attested at several different stages of the cycle in a given language family. So many solid examples exist of passage from one "contiguous" stage to another, that it seems logical to believe in the reality of the entire cycle itself.

Most of these interrelationships are bidirectional, as indicated by the arrows in the diagram. We will briefly take them up one by one.

2.1 *Contractive*: from complex-consonant to simple-consonant monosyllables.

The documented history of many languages (including Chinese, Tibetan, and Burmese) shows sometimes radical simplification of an earlier consonantism: e.g. Old Chinese **tsywet** 'emerge' > Mandarin **chū**; Written Tibetan **brgyad** 'eight' > Mod. Lhasa **cɛ**; Written Burmese **krwat** 'leech' > Mod. Rangoon **cwa?**. This has been carried to an extreme in branches of the family like Loloish, where the rich syllable canon set up for PTB *[see above]* has been eroded, e.g. in Lahu to

$$(C) \quad \overset{T}{V} \quad ,$$

as in 'snake' WB **mrwe**, Lh. **vɨ**; 'rat' WB **krwak**, Lh. **fâ?**; 'eight' WB **hrac**, Lh. **hí**. Note, however, that compensatory mechanisms -- proliferation of vowels and/or tones *[see above 1(h)]* -- usually operate to counteract the consonantal impoverishment.[15]

2.2 *Expansive*: from complex monosyllables to sesquisyllables.

This development has nothing characteristically SE Asian about it, since it is basically the low-level phonetic phenomenon of *epenthesis* that is at work -- the breaking up of hard-to-pronounce consonant clusters by shwa-insertion.   It seems reasonable to assume that many of the orthographic clusters of a language like Written Tibetan were broken up in actual speech by such epenthetic vowels, so that e.g. **brgyad** 'eight' must have been pronounced something like [brəgyat].[16]

2.3 *Expansive*: from monosyllables to dissyllables

Radical consonantal simplification can pose a serious problem for monosyllabic languages in terms of *pernicious homophony.*   Keeping potentially homophonous morphemes distinct requires adding some "phonological bulk" or redundancy.   One way to achieve this is to increase the number of syllables per word. Occasionally this is done by meaningless "mechanical" extrusions of the monosyllable's phonetic substance (2.31); usually, however, the strategy is to create dissyllabic compounds or collocations where each of the constituent syllables is fully meaningful (2.32).

2.31 Dissyllables without morpheme boundary.

Some TB languages have developed "echo vowels" which serve to resyllabify an original monosyllable into a 2-syllable sequence.   The PTB root ***p-wak** 'pig' has become the dissyllabic form **wo-ko** in Lotha Naga.   Similarly, PTB ***lam** 'road' has developed into **la-ma** in Kokborok (Bodo-Garo group) and to **lə-mu** in Khoa (Abor-Miri-Dafla group).   In Garo, a medial vowel is sometimes split into two morae separated by glottal stop: PTB ***g-sat** 'kill" > Garo **soʔot**.

2.32 Dissyllables with morpheme boundary: compounding and affixation.

Much more important as a strategy for providing compensatory phonological bulk is the process of *compounding.* Many dialects of Southern American English have merged the rhymes /-in/ and /-en/, rendering pairs like *pin / pen*

homophonous. To disambiguate such cases, these dialects have introduced compounds like *stick-pin / ink-pin*. East and Southeast Asian languages have proceeded in exactly analogous fashion.

Chinese is an excellent case in point. Classical Chinese, with its relatively rich consonantism, was strictly monosyllabic, with the syntactic word and the phonological syllable virtually coextensive. In phonologically eroded modern dialects like Mandarin, however, the vast majority of *words* are now dissyllabic, though almost all of them are still analyzable into monosyllabic constituent morphemes.[17]

In Loloish also, the consonantal simplification of monosyllables has led to homophony on a grand scale. There are e.g. at least five Lahu morphemes pronounced **ha**[18] which descend from once consonantally distinct PTB etyma:

	PTB	PLB	WB	Lahu monosyllable	Lahu dissyllabic collocations
'hundred'	*b-r-gya	*ʔra¹	ra	ha	tê ha
'moon'	*s-gla	*s-la³	la'	ha	ha-pa
'tongue'	*s-lya	*s-l(y)a¹	hlya	ha	ha-tɛ̄
'spirit, likeness, shadow'	*s-hla	*sla³	hla' 'beautiful'	ha	ɔ̀-ha
'winnow'	*g-ya(:p)	*ʔ-ya¹	Akha zá	ha	ha ve

[**ha** 'hundred' is not usable by itself, but functions as a 'round-number classifier' which must always be preceded by a numeral (e.g. **tê ha** 'one hundred'); the **-pa** in 'moon' is a meaningless noun suffix, ubiquitous in TB (cf. WT **zla-ba** 'moon')[19]; the -tɛ̄ in 'tongue' looks like it once had an independent meaning, but recurs nowhere else in the language; the **ɔ̀-** prefix in 'spirit' (< PTB ***ʔaŋ-**) occurs as a bulk-provider before hundreds of Lahu roots; the particle **ve** in **ha ve** 'to winnow' is a nominalizer that occurs in the citation form of verbs (much like Eng. *to*), serving to distinguish verbs from any homophonous nouns.][20]

2.4 *Contractive*: **from dissyllables to monosyllables.**

Once we have a dissyllabic structure in close phonological juncture, the stage is set for the reverse swing of the cyclical pendulum: the fusion of the two syllables into one. Several subtypes of "trans-syllabic absorption" may be distinguished:

Sometimes the two syllables have been so thoroughly fused that they can only be teased apart by comparative evidence. I have explained the peculiar Angami Naga form **pfhə** 'bitter', where the labiodental initial stands out sharply against the simple velar attested everywhere else in TB (< ***ka**), in terms of the fusion of an original dissyllable ***ka-ba** (where **-ba** was a nominalizing citation particle):

<div align="center">

*ka-ba ⟶ *ka-wa ⟶ *kwa ⟶ pfhə.[21]

</div>

The Garo word for 'I; me' is **aŋ**, whereas most other TB languages have forms which directly reflect ***ŋa**. Metathesis will not do as an explanation! It seems more likely that the well-attested TB pronominal prefix ***a-** was combined with this morpheme, after which the final (originally root-)vowel was dropped:

<div align="center">

*ŋa ⟶ *a-ŋa ⟶ aŋ

</div>

Speaking of pronominal fusions, the Mandarin pluralizing suffix **-men**, used only with pronouns and a few nouns referring to human beings, has lost its rhyme and fused its initial with the vowel of the preceding pronoun in rapid colloquial: **wŏ-men** 'we' > **wŏm**, **ni-men** 'you pl.' > **nim**, **tā-men** 'they' > **tām**. This has incidentally had the effect of marginally reintroducing final **-m** into the Mandarin syllable canon, from which it had disappeared centuries ago by merging with final **-n** (cf. Cantonese **saam**, Mand. **sān** 'three').

New data on the Qiangic languages (TB of Sichuan) shows that the Northern dialects have a strong tendency to apocopate the vowel of the 2nd elements of compounds, leading to secondary monosyllables with final consonants (e.g. voiced spirants) that are sometimes highly untypical of TB:

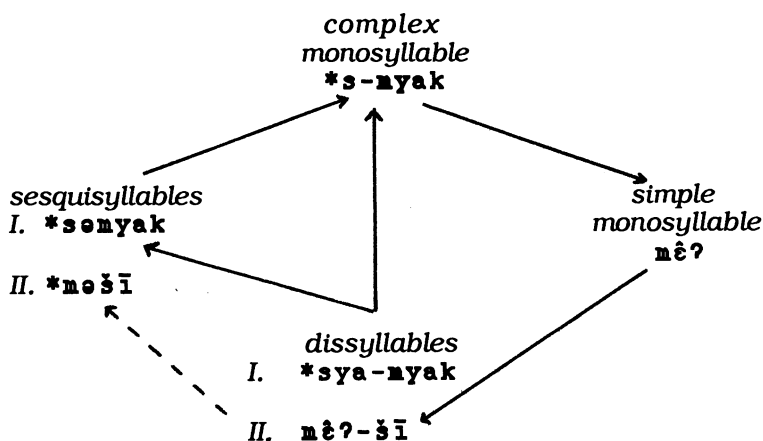|	S.Qiang (Taoping)	N.Qiang (Mawo)
'seed'	zuə-pə	zəp
'day after tomorrow'	zuə-za	tśhaz
'stove'	sy-dy	syt
'fifteen'	xa-ŋa	haŋ

When the 2nd syllable in a collocation begins with a vowel, and when its meaning is grammaticalized or abstract (which facilitates destressing), it is susceptible of being absorbed into the vocalic nucleus of a preceding open syllable. This often leads to a "bulging" of the vowel until it is diphthongal, or at least "sesquimoral" -- a mora and a half long.[22]

2.5 *Contractive*: **from dissyllables to sesquisyllables.**

Perhaps the most interesting stage in the cycle is the one where we can catch a dissyllabic compound in the process of being reduced to a sesquisyllable, but before the semantic content has quite disappeared from the unstressed minor syllable -- or at any rate while it may still be deduced from comparative data.[23] Much as the heart of a chivalrous knight would be quickened by hapless damsels in days of yore, so might we well be moved by the plight of these poor "syllables in de-stress". Let us quickly chase a few of them around the cycle:
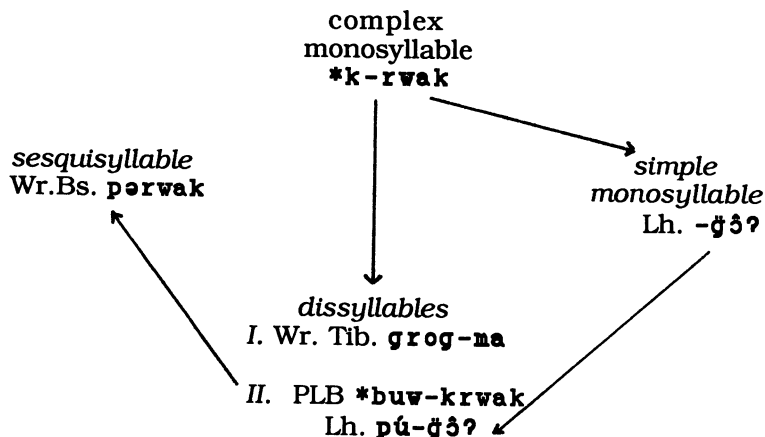
2.51 *EYE*

Pre-PTB *sya-myak > PTB/PLB *səmyak or *s-myak > Lahu mɛ̂ʔ > Lahu mɛ̂ʔ-ší > hypothetical future *məší

Here the only directly attested forms are the Lahu ones, yet there is good evidence for all the other stages. The high

stopped tone of **nê?** indicates a PLB ***s-** prefix before the nasal [Matisoff 1972, pp. 24, 58-61]. This prefix with bodypart terms is generally explained as deriving from the full morpheme ***sya** 'flesh; animal' [Benedict 1972, p. 106]. It may be supposed that the dissyllable ***sya-nyak** passed through a sesquisyllabic stage ***sənyak** during the "prefixization" process. Meanwhile, the monosyllabic form **nê?** is replaced in most contexts by the dissyllabic compound **nê?-ší** in modern Lahu, where the second element means 'round object'. We may confidently expect that in several hundred years this compound will itself degrade into a new sesquisyllable, perhaps ***nəší**, as the carousel goes around once again.

2.52 ANT

complex
monosyllable
*k-rwak

sesquisyllable
Wr.Bs. **pərwak**

simple
monosyllable
Lh. **-ɠɔ́?**

dissyllables
I. Wr. Tib. **grog-ma**

II. PLB ***buw-krwak**
Lh. **pú-ɠɔ́?**

The first syllable of the Wr.Tib. form directly reflects the PTB prototype with prefixal velar (one of the generally accepted "animal prefixes" of TB). The high-stopped tone of the Lahu syllable **-ɠɔ́?** also reflects the ***k-** prefix [Matisoff 1972, pp. 68-70], although this syllable does not occur independently in modern Lahu. Instead we find the compound **pú-ɠɔ́?**, where the first element is definitely a sandhi variant of the morpheme **pû** 'bug' (< PTB ***bəw**), which occurs in a large number of insect names. This compound formation, innovative with respect to TB as a whole, must be referred back to the PLB stage, since a sesquisyllable with this same element is directly attested in Burmese, both in the orthography and in the modern pronunciation /pəywɛ?/.

2.53 SON-IN-LAW
Wr.Bs. **səmak** / *Wr.Tib.* **mag-pa**, *Dhimal* **hma-wa**

sesquisyllable

*simple
monosyllable*

Wr.Bs. **səmak**

*ma:k

dissyllables
Wr.Tib. **mag-pa**

PLB ***za-mak(-pa)**

PLB ***ʔaŋ-mak(-pa)**
Lahu **ɔ̀-má(-pā)**

Here the proto-monosyllable ***ma:k** was quite simple to begin with, so that at an early stage it was compounded with various morphemes in the daughter languages, especially with the suffix **-pa** (which after nouns referring to humans functioned sometimes as a masculine morpheme contrasting with **-ma** 'feminine'). The minor syllable of the directly attested sesquisyllabic form in Burmese seems obviously to be a reduction of the full morpheme ***za** 'son; child'. This time the Lahu forms do not reflect the same formation as in Burmese, but only the meaningless prefix ***ʔaŋ-**, with the masculine suffix **-pā** sometimes thrown in for additional redundancy.

2.54 LUNGS

*complex
monosyllable*
Lushai **tśuap**

sesquisyllable
*tšəwap

*simple
monosyllables*
*tsi 'lung'
*-wap 'spongy'

dissyllables
pre-Lushai ***tsi-wap**

Jingpho **sìn-wóp**

pre-Lahu ***ʔaŋ-tsi(-pwap)**
Lahu **ɔ̀-chî(-phô?)**

I have demonstrated at length that the complex monosyllabic form for 'lungs' in Lushai (Kuki-Chin branch of TB), t**śuap**, actually derives from a dissyllabic prototype, ***tsi-wap**, where the 2nd element is an adjectival root meaning 'spongy; porous'.[24] This is made clear by the Jingpho compound **sin-wóp** 'lungs' ("spongy liver"), where the first element derives from PTB ***m-sin** 'liver', and which is paradigmatically opposed to the compound **sin-jà?** 'liver' (lit. "solid liver"). The 2nd syllable of the Lahu form reflects ***tsi**, while the 3rd syllable might derive from a prefixed variant of the morpheme for 'spongy', ***p-wap**.[25]

2.6 *Contraction*: from sesquisyllables to monosyllables

Finally, we may close the circle by noting that there are multiple paths by which a sesquisyllable may be reduced to a monosyllable.

Most commonly perhaps, the minor syllable is simply lost by procope. This is what happened regularly in the history of Vietnamese, once a sesquisyllabic Mon-Khmer language that became monosyllabic and highly tonal under massive and prolonged Chinese influence.

Alternatively, the shwa vocalism of the pre-syllable may disappear by syncope, creating a complex monosyllable, which is probably the scenario leading to the Lushai form for 'lungs' just discussed.

Still another possibility is a development I have called *prefix preemption*,[26] which occurs especially when the initial of the major syllable is a "weak" consonant (a liquid, semivowel, or nasal). In these cases the initial of the prefixal or "minor" syllable may drive it out entirely. Many examples may be given from TB (cf. Maru **bit** 'four' < PTB ***b-ləy**, cited in I(g), above), and the same phenomenon may be noted in colloquial Siamese:

> **məlèt** 'seed' (formal) > **nèt** (colloq.)
> **məleɛŋ** 'bug' (formal) > **neɛŋ** (colloq.)

3. L'économie de la syllabe: decay and rebirth.

In conclusion, let me repeat a passage with which I ended another recent discussion of this general topic:

"The monosyllabic languages of East and Southeast Asia show an uncanny homeostatic ability to regulate themselves in cyclic swings of expansion and contraction. What is absorbed and incorporated *here* will be diffused or extruded *there*.

"The accretional or augmentative tendencies do not of course stand in a simple one-to-one replacement relationship versus the tendencies toward reduction and attrition. Things are more indirect and slow-moving than that. Nonetheless, it is hard not to believe in some kind of overarching regulatory principle which eventually ensures that things will not go too far in any one direction. There is no harm in referring to this by some functional label like 'the economy of the syllable'.

"In a more cosmic vein, these phenomena furnish one more bit of reassuring evidence that the forces of creativity have nothing to fear from the forces of destruction."[27]

NOTES

[1]A version of this paper was presented at the 15th annual meeting of the Berkeley Linguistics Society in February 1989. My thanks to John B. Lowe for help in formatting the present version.Some of this material is based upon work done at the Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) Project, supported by the National Science Foundation under Grant No. BNS-867726 and by the Division of Research Programs of the National Endowment for the Humanities, an independent federal agency, under Grant No. RT-20789-87.

[2]These monosyllabic families include Sino-Tibetan/Tibeto-Burman, Tai-Kadai, Hmong-Mien (Miao-Yao), and the Viet-Muong branch of Austroasiatic (Mon-Khmer). The rest of Austroasiastic is largely "sesquisyllabic" (comprising morphemes "one syllable and a half" in length")*[see below]*. Only Austronesian and Japanese contain large numbers of truly dissyllabic morphemes.

[3]Matisoff 1973, p. 78. Henderson (1985) later dubbed this phenomenon "feature shuffling."

[4]There seems to be a rough inverse correlation between the typical number of syllables in a language's morphemes and the phonetic complexity of the individual syllables.

[5]These are elegantly discussed in Diffloth 1973, though I alone am responsible for coining the ugly term "incopyfix."

[6]Pace Miller 1958.

[7]When prefixes were present we may suppose that they were often vocalized by a following shwa, so that such syllables were phonetically "sesquisyllabic" [below].

[8]We are using the symbol ">" to indicate the direction of influence. "PLB" is Proto-Lolo-Burmese, one of the major subgroups of PTB. Lahu is a member of the Central Loloish branch of Lolo-Burmese.

[9]The traditional term "rhyme" refers to the entire syllable except for the initial consonant. The "glides" /w y r l/ are inherently ambiguous with respect to this concept, sometimes functioning as if they were part of the initial but sometimes as if part of the vocalic nucleus. Thus,

PLB *wak 'pig' > Lh. vì?, but PLB *twak 'emerge' > Lh. tɔ̂?

PLB *s-ya¹ 'winnow' > Lh. ha, but PLB *hya¹ 'swidden' > Lh. hɛ.

[10]Matisoff 1978b, pp. 22-4. All examples found so far reconstruct with either *s- or *?-, two prefixes which are especially similar in their diachronic behavior in Lolo-Burmese.

[11]These rhymes were reconstructed as *-iy and *-uw in the original version of Benedict 1972, but were later reinterpreted as *-əy and *-əw. Under either analysis they are functionally equivalent to bimoraic vowels.

An analogous process is posited for the development of the Old Chinese final semivowels *-w and *-y into the Middle Chinese final voiced stops *-g and *-d. See Benedict 1948.

[12]The term *tonogenesis* was first used in Matisoff 1970, 1973, though the phenomenon itself had been explained long before by Haudricourt (1954).

[13]Diffloth [p.c.] reports that the Chong language (of the Pearic branch of Mon-Khmer, spoken in Cambodia) has a 4-way contrast among clear / breathy / creaky-breathy / and "forced" (i.e. tight and faucalized) phonation types.

As an adjunct or alternative to "registrogenesis" many Mon-Khmer languages have developed highly complex vocalic systems to compensate for consonantal mergers. Khmer itself has 31 vocalic nuclei (including some triphthongs), while Bru (Katuic group, Vietnam) has no fewer than 41.

[14]The term *sesquisyllable* was introduced in Matisoff 1973 to refer to words that are "a syllable and a half" in length. Sesquisyllables consist of a fully stressed "major syllable" preceded by an unstressed "minor syllable" that usually has shwa-vocalism (e.g. **phənòm, rəbiap, məlèt**). This sort of syllable structure is especially characteristic of the Mon-Khmer family (except for Vietnamese, which has become strictly monosyllabic under Chinese influence), though it is also widespread in Tibeto-Burman and occasionally encountered in Tai-Kadai.

[15]In addition to these phonemic compensations (transphonologizations) resulting from regular sound change, many Lahu syllables are more complex phonetically than this bare-bones canon would suggest. Loanwords, affective vocabulary, and morphophonemic processes like vocalic fusion combine to reintroduce such features as secondary nasalization, glottalization, labial glides, long vowels, and diphthongs into the language. Any of these historically secondary features might acquire considerable importance in the future. "Once some feature is present phonetically in a SE Asian monosyllable, no matter how

redundant or trivial it may appear, it is available for future exploitation and transphonologization" [see Matisoff 1989a].

[16]These syllables are crucially different from those much more interesting sesquisyllables to be considered below (2.5), where the minor syllable derives from a recognizable separate morpheme.

[17]In a discussion of homophony in an encyclopedia article on Chinese, Y.R. Chao once concocted three little stories in Classical Chinese style consisting entirely of monosyllabic words pronounced homophonously in Mandarin (**shí, jí,** and **yí**, respectively, under various tones). For the Old Chinese listener, these stories would have been understandable even orally, since most of the syllables were still pronounced differently. Thus the three words of the title of the story *Ten Stone Lions* (Mand.*Shí Shí Shí*) were pronounced something like **Dyəp Dyak Syər* in OC. For the stories to be auditorily understandable to a modern Mandarin speaker, they would have to be recast using dissyllabic compounds or collocations to differentiate the now individually homophonous syllables (e.g. **shí-ge** 'ten', **shí-tou** 'stone', **shī-zi** 'lion').

[18]Besides sharing the same initial and vowel, these syllables are also *tonally* homophonous, all being under the mid-tone (unmarked in the transcription). Many other **ha**-morphemes occur under the other 6 tones!

[19]A homophonous suffix -**pa/-ba** occurred as a nominalizer after verbal roots. See the discussion of the Angami form for 'bitter', below 2.4.

[20]Cf. also **ha-kə** 'small winnowing tray', **ha-ma(-qō)** 'large winnowing sieve'. It should be noted that many TB compounds are 3, 4, or even more syllables long, as in the "long form" of the word for 'large winnowing sieve', **ha-ma-qō**, where the 2nd syllable is a recurrent but meaningless suffix and the 3rd syllable means 'concave object'.

[21]See Matisoff 1982, p. 24. The development of labiodentals from labiovelars is well attested elsewhere in TB, e.g. PTB **kwəy* ' dog' > Lh. **phí** *[see above 1(b)]*.

[22]I have recently discussed this topic at length elsewhere (Matisoff 1989a, 1989b). Falling into this category are the innumerable syllables of Pekinese Mandarin with retroflex vowels that derive from a diminutive suffix that is itself an unstressed variant of the full morpheme **ér** 'child', e.g. **niăor** 'birdie', **xiăo-háir** 'child', **guŏ-tiěr** 'potsticker dumpling'.

[23] The loss of morphemic identity from one or both of the constituents of an original compound is a danger which every language faces, with many familiar examples citable from the history of English: *hussy* (< HOUSE + WIFE), *bonfire* (< BONE + FIRE), *nostril* (< NOSE + THYREL [obs.] 'hole'), *window* (< WIND + EYE), *daisy* (< DAY's + EYE), etc.

[24]See Matisoff 1978a, pp. 113-23.

[25]The examples given so far are all from TB, though similar cases of sesquisyllabization of compounds can easily be found in e.g. Tai-Kadai. Thus the minor syllable of Siamese **sədaa** 'navel' is derivable from **siaj** 'line; band' via the umbilical cord, while the unstressed **ma-** in many names of fruits and vegetables (e.g. **mamûaŋ** 'mango', **maphráaw** 'coconut', **makhŷa** 'eggplant') is a reduction of PTai **hmaak* 'fruit'. Cf. Li 1977, pp. 92, 75.

[26]See Matisoff 1979, p. 24.
[27]Matisoff 1989a, p. 185.

REFERENCES

Benedict, Paul K. 1948. "Archaic Chinese *g and *d." Harvard: Journal of Asiatic Studies 6, 313-37.

_____. 1972. *Sino-Tibetan: a Conspectus*. Contributing editor, James A. Matisoff. Cambridge: Cambridge University Press.

Diffloth, Gérard. 1973. "Minor syllable vocalism in Senoic languages." In Philip N. Jenner, et al., eds. *Austroasiatic. Studies, Part I*. (Pacific Linguistics, Spec. Pub. #13.). Honolulu: University of Hawaii Press, 229-47.

Haudricourt, André-Georges. 1954. "De l'origine des tons en viétnamien." Journal Asiatique 242.68-82.

Henderson, Eugénie J.A. 1985. "Feature shuffling in Southeast Asian languages, or How big is a segment?" In S. Ratanakul, D. Thomas, and S. Premsirat, eds., *Southeast Asian Linguistic Studies Presented to André G. Haudricourt*, pp. 1-22. Bangkok, Mahidol University.

Li Fang-kuei. 1977. *A Handbook of Comparative Tai*. Honolulu: University of Hawaii Press.

Matisoff, James A. 1970. "Glottal dissimilation and the Lahu high-rising tone: a tonogenetic case-study." Journal of the American Oriental Society 90.1, 13-44.

_____. 1972. *The Loloish Tonal Split Revisited*. Research Monograph #7, Center for South and Southeast Asia Studies, University of California, Berkeley.

_____. 1973. "Tonogenesis in Southeast Asia." In L.M. Hyman, ed., *Consonant Types and Tone*. (Southern California. Occasional Papers in Linguistics #1.) Berkeley and Los Angeles, University of California Press, 71-96.

_____. 1978a. *Variational Semantics in Tibeto-Burman*. Philadelphia: ISHI Publications.

_____. 1978b. "Mpi and Lolo-Burmese microlinguistics." Monumenta Serindica (Tokyo), #4. 36 pp.

_____. 1979. "Problems and progress in Lolo-Burmese: Quo Vadimus?" Linguistics of the Tibeto-Burman Area 4.2, 11-43.

_____. 1982. "Proto-languages and proto-Sprach-gefühl." Linguistics of the Tibeto-Burman Area 6.2, 1-64.

_____. 1989a. "The bulging monosyllable, or *The mora. the merrier*: echo-vowel adverbialization in Lahu." In J.H.C.S. Davidson, ed., *South-East Asian Linguistics: Essays in honour of Eugénie J.A. Henderson*. London: School of Oriental and African Studies, 163-97.

_____. 1989b. "The three palatal suffixes of Tibeto-Burman." Talk presented at 22nd International Conference on Sino-Tibetan Languages and Linguistics, University of Hawaii, Honolulu (October 7).

_____. 1989c. "Tone, intonation, and sound symbolism in Lahu: loading the syllable canon." Linguistics of the Tibeto-Burman Area 12.2, 147-63.

Miller, Roy Andrew. 1958. "The Tibeto-Burman infix system." Journal of the American Oriental Society 78.3.

BERKELEY LINGUISTICS SOCIETY

2337 Dwinelle Hall, University of California, Berkeley, CA 94720 USA
Telephone: (415) 642-5808, (415) 642-2757; e-mail: bls@garnet.berkeley.edu

The Berkeley Linguistics Society is a graduate student organization at the University of California, Berkeley. The Society holds its annual meeting during Presidents' Day weekend in the middle of February. The meeting consists of three parts: a General Session on topics of varied linguistic interest, a Parasession on a particular linguistic topic, and, beginning in 1990, a Special Session on general linguistic topics of a certain geographic area. Around the middle of September, a Call-for-Papers is sent to those on our mailing list, and abstracts of papers are due around the middle of November. After reviewing and selecting the abstracts, the Society sends out the conference program in early January. The Proceedings of the conference are published yearly by the Society and become available in the fall. Order forms for proceedings are mailed with the Call-for-Papers and can be obtained from the above address. All orders require prepayment.

Book Prices

	Parasession Topics	Individuals	Institutions
BLS-1 (1975)		$7.00	$10.00
BLS-2 (1976)		$6.00	$10.00
BLS-3 (1977)		$6.00	$10.00
BLS-4 (1978)		$6.00	$10.00
BLS-5 (1979)		$6.00	$10.00
BLS-6 (1980)		$6.00	$10.00
BLS-7 (1981)		$6.00	$10.00
BLS-8 (1982)		$10.00	$14.00
BLS-9 (1983)		$6.00	$10.00
BLS-10 (1984)	Subordination	$10.00	$14.00
BLS-11 (1985)	Poetics, Metrics, and Prosody	$10.00	$14.00
BLS-12 (1986)	Semantic Typology	$10.00	$14.00
BLS-13 (1987)	Grammar and Cognition	$11.00	$15.00
BLS-14 (1988)	Grammaticalization	$12.00	$16.00
BLS-15 (1989)	Theoretical Issues in Language Reconstruction	$12.00	$16.00
BLS-16 (1990)	Legacy of Grice	$16.00	$20.00
BLS-16S (1990)	Special Session on General Topics in American Indian Linguistics	$8.00	$12.00

California Sales Tax

7.0% for residents of Alameda, Contra Costa, San Diego, San Francisco, San Mateo, and
       Santa Clara counties
6.5% for residents of Fresno, Imperial, Inyo, Los Angeles, Madera, Monterey, Riverside,
       Sacramento, San Benito, San Bernardino, Santa Barbara, and Santa Cruz counties
6.0% for all other California residents

Shipping and Handling

	Within USA	Outside USA
BLS 1,2,6,7,9,11,12,13,14,15,16S	$2.00	$3.00
BLS 3,4,5,8,10,16	$2.50	$4.00

ORDERS MUST BE ACCOMPANIED BY PAYMENT
Make checks and money orders payable to the Berkeley Linguistics Society