

Proceedings of the Twenty-Seventh Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Gesture and Language (2001)

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

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 TWENTY-SEVENTH ANNUAL MEETING OF THE
BERKELEY LINGUISTICS SOCIETY**

February 16-18, 2001

GENERAL SESSION

and

PARASESSION on GESTURE AND LANGUAGE

Edited by

Charles Chang, Michael J. Houser, Yuni Kim, David Mortensen,
Mischa Park-Doob, and Maziar Toosarvandani

**Berkeley Linguistics Society
Berkeley, CA, USA**

Berkeley Linguistics Society
University of California, Berkeley
Department of Linguistics
1203 Dwinelle Hall
Berkeley, CA 94720-2650
USA

All papers copyright © 2005 by the Berkeley Linguistics Society, Inc.

All rights reserved.

ISSN 0363-2946

LCCN 76-640143

Printed by Sheridan Books
100 N. Staebler Road
Ann Arbor, MI 48103

TABLE OF CONTENTS

A note regarding the contents of this volume	vi
Foreword	vii

GENERAL SESSION

Vowel Harmony and Cyclicity in Eastern Nilotic	1
ERIC BAKOVIC	
Constructional Effects of <i>Just Because ... Doesn't Mean ...</i>	13
EMILY M. BENDER and ANDREAS KATHOL	
Modification of the Noun in Somali: Comparative Evidence from other Cushitic Languages	27
SABRINA BENDJABALLAH and PATRICIA CABREDO HOFHERR	
Syntactic Ergativity in Light Verb Complements	39
BALTHASAR BICKEL and JOHANNA NICHOLS	
The Intersubjectivity of 'Self' and 'Other' Tokens in Communication: Moving Between Represented Subjectivity and Conceptual Objectivity	53
MELINDA Y. CHEN	
Voiceless Tone Depressors in Nambya and Botswana Kalang'a	65
LAURA J. DOWNING and BRYAN GICK	
Object Marking and Agentivity in Navajo Causatives	81
SUZANNE GESSNER	
An Account of Compensatory Lengthening without Moras	95
BEN HERMANS	
Interactions between Constructions and Constraints in VP Ellipsis and VP Fronting	107
JONG-BOK KIM	
Relexification: A Process Available to Human Cognition	125
CLAIRE LEFEBVRE	

Contact, Attrition, and Structural Shift: The Case of Oroqen	141
FENGXIANG LI	
Cross-Language Perception of Syllable Affiliation: Effects of Voicing and Language Background	157
BYUNG-JIN LIM, KEN DE JONG, and KYOKO NAGAO	
Speech Reporting Practices and Subjectivities: Evidence from an Amazonian Society	169
LEV MICHAEL	
Lexical Tone and Markedness in Standard Thai	181
BRUCE MORÉN and ELIZABETH ZSIGA	
The Relationship between the Source Characteristic and Frequency	193
HANSANG PARK	
A Unified Account of Essentially Contested Concepts	205
JASON D. PATENT	
Direct Quotation in Japanese: A Site for Stereotyping	215
KATSUE AKIBA REYNOLDS	
Valence, Voice, and Noun Incorporation in Karajá	229
EDUARDO RIVAIL RIBEIRO	
Highly-articulated Representations in Early Grammars: Evidence from English and French	245
YVAN ROSE	
On the Phonologically Driven Non-Realization of Function Words	257
ELISABETH SELKIRK	
Spatial Structuring in Spoken and Signed Language	271
LEONARD TALMY	
Pronoun Borrowing	301
SARAH G. THOMASON and DANIEL L. EVERETT	
Courtroom Accommodation: Lawyers' Use of Foreigner Talk	317
ELIZABETH CLOSS TRAUGOTT	
Discourse Pragmatics vs. Prescriptive Stylistics	327
OLGA T. YOKOYAMA	

PARASESSION: GESTURE AND LANGUAGE

Indicating Participants in Actions: From Prelinguistic Gestures to Signed Languages	341
SHANNON CASEY	
Co-expressivity of Speech and Gesture: Manner of Motion in Spanish, English, and Chinese	353
SUSAN DUNCAN	
The Development of Gesture, Speech, and Action as Communicative Strategies	371
BARBARA F. KELLY	
An Analysis of Shifts in Participation Roles in Japanese Storytelling in Terms of Prosody, Gaze, and Body Movements	381
CHISATO KOIKE	
Gesture, Information Flow, and NP Form	393
LORRAINE E. KUMPF	
Grammar and Gesture in American Sign Language: Implications for Constructing Meaning	405
SCOTT K. LIDDELL	
Prosodic and Gestural Cues for Navigations around Mental Space	421
JUNE LUCHJENBROERS	
Does Gesture Help Processes of Speech Production? Evidence for Conceptual Level Facilitation	435
ALISSA MELINGER and SOTARO KITA	
What Do Speech-Gesture Mismatches Reveal about Speech and Gesture Integration? A Comparison of English and Turkish	449
ASLI ÖZYÜREK	
Gestures and Self-Monitoring in Speech Production	457
MANDANA SEYFEDDINIPUR and SOTARO KITA	
Comparing the Informational Content of Speech/Gesture and Sign Language	465
SARAH TAUB and PILAR PIÑAR	

A note regarding the contents of this volume

The following authors also presented papers at the conference, though their work does not appear in this volume: Abbas Benmamoun, Susan Goldin-Meadow, Samuel H. Hawk, Kristin Homer & Laura Michaelis, and Maria A. Yakubovich.

The paper included here by Bendjaballah & Cabredo Hofherr was originally written for the Special Session volume on Afroasiatic Linguistics, but was inadvertently left out of that volume.

Foreword

The editors would like to thank the contributors and the readership for their everlasting patience (and good humor) during the several years it has taken for this volume to come to press. On behalf of BLS and quite a few generations of officers, we hope you will accept our humble apologies for the delay, and that you find the volume to be of value.

Charles Chang, Michael J. Houser, Yuni Kim, David Mortensen,
Mischa Park-Doob, and Maziar Toosarvandani

Volume editors

GENERAL SESSION

Vowel Harmony and Cyclicity in Eastern Nilotic*

ERIC BAKOVIĆ

University of California, San Diego

0. Abstract

Vowel harmony in the Eastern Nilotic languages Maasai (Tucker & Mpaayei 1955) and Turkana (Dimmendaal 1983) is dominant-recessive: a [+ATR] vowel in either a root or a suffix causes all other vowels in the word to become [+ATR]. The phonemic low vowel in both of these languages behaves differently depending on its position in the word relative to the source of [+ATR], a fact that has previously been accounted for by two directional [+ATR] spreading rules subject to distinct conditions (Archangeli & Pulleyblank 1994, Albert 1995). I propose instead that the distinction between these two directions of harmony is (indirectly) determined by the cycle, and that harmony is due to a single bidirectional mechanism. A cyclic account makes a number of nontrivial and restrictive predictions that a directional account does not. First, a grammar in which the conditions on the two directions of harmony are somehow reversed is predicted not to exist. Second, the cyclic account crucially depends on the indirect blocking of harmony by another process that respects the cycle, and so the absence of such a process entails the absence of a directional asymmetry. Third, a grammar in which harmony only operates in one direction and not the other is predicted not to exist. Finally, only the cyclic account can readily handle a set of additional facts in Turkana.

1. Background: Stem Control vs. Dominance¹

Research on vowel harmony has revealed that there are two basic types of vowel harmony systems, stem-controlled and dominant-recessive. In stem-controlled systems, the harmonic feature value of an affix vowel is dependent on the harmonic feature value of the adjacent vowel in the stem to which the affix is at-

* I thank Sharon Rose, Colin Wilson, and audiences at BLS, UCLA, and UCI for helpful comments on some of the contents of this paper. This work is an outgrowth of my Rutgers University dissertation (Baković 2000), so those thanked there are also thanked here. Errors are mine.

¹ Here and throughout, ‘stem’ refers to any morphological constituent to which an affix may attach; ‘root’ refers to the innermost such constituent, the ultimate stem of affixation. Where the distinction between them is irrelevant, I refer to the ‘root/stem’.

tached. The vowel of the affix closest to the root changes to agree with the adjacent root vowel, the vowel of the next closest affix changes to agree with the adjacent vowel of the closest affix, and so on. Some examples are given in (1).²

(1) Stem-controlled vowel harmony

Akan (Niger-Congo, Kwa; Schachter & Fromkin 1968)

- a. /E + √bu + O/ → [ebuo
 CLASS + nest + SFX ‘nest’
- b. /E + √bu + O/ → [ɛbuo
 CLASS + stone + SFX ‘stone’

Tangale (Afro-Asiatic, W. Chadic; Kidida 1985)

- c. /√tug + O/ → [tugo
 pound + NOM ‘pounding’
- d. /√wud + O/ → [wudo
 farm + NOM ‘farming’

In dominant-recessive systems, the harmonic feature value of all vowels in the domain of harmony (here, the word) is dependent on whether or not one of them is underlyingly specified for the ‘dominant’ value of the harmonic feature, [αhf]. If any vowel (stem or affix) is [αhf], then all vowels surface as [αhf]; if all are [$-\alpha hf$] (‘recessive’), then all surface unchanged as [$-\alpha hf$]. Examples from the two Eastern Nilotic languages that are the focus of this paper are given in (2).

(2) Dominant-recessive vowel harmony

Maasai (Tucker & Mpaayei 1955, Archangeli & Pulleyblank 1994)

- a. /kI + √norr + u/ → [kinorru
 1PL + love + EF ‘we shall love’
- b. /√isuj + Iʃɔ + re/ → [isujiʃore]
 wash + INTRANS + APPL ‘wash with something!’

Turkana (Dimmendaal 1983, Albert 1995)

- c. /ɛ + √los + I/ → [elosi]
 3 + go + ASP ‘s/he will go’
- d. /ɛ + √kokɔ + un + I + o/ → [ekokounio]
 3 + steal + VEN + IMP + PL.PASS ‘they are being stolen’

In Akan and Tangale (1), root/stem vowels are constant while affix vowels vary in terms of the harmonic feature [$\pm ATR$]. In Maasai and Turkana (2), both stem and affix vowels potentially vary in terms of the harmonic feature [$\pm ATR$], depending on whether or not any other vowel in the word underlyingly bears the dominant

² In these and other examples, the radical symbol ‘√’ indicates the root morpheme, and capital letters represent vowels whose underlying specification for the harmonic feature cannot be determined. Underlining in underlying forms indicates the vowel instigating harmony, and in surface forms it indicates the propagation of the harmonic feature throughout the word.

value [+ATR]: if so, then all vowels change if necessary to surface as [+ATR]; if not, all vowels surface with their underlying recessive value [−ATR].

I have argued elsewhere (Baković 2000) that the fundamental difference between stem-controlled harmony and dominant-recessive harmony is one of *cyclicity*, as defined in (3) below. (Note that this definition is purposely agnostic as to the particular theoretical mechanism that is employed to account for it.)

- (3) A phonological process is *cyclic* iff it systematically fails to apply to stems of affixation.

Stem-controlled harmony processes are cyclic because disharmony between stem and affix vowels is systematically resolved by changing the harmonic feature value of the affix vowels, not the stem vowels. On the other hand, dominant-recessive harmony processes are *noncyclic*, because disharmony between stem vowels and affix vowels is resolved sometimes by changing affix vowels and sometimes by changing stem vowels. In a dominant-recessive harmony system, vowels with the dominant feature value are the instigators of harmony; in a stem-controlled harmony system, cyclicity determines the instigator.

The cyclic approach to stem-controlled vowel harmony has a number of advantages over the familiar standard approach involving processes of directional feature propagation, two of which I briefly point out here. (For further arguments for and elaboration of the cyclic approach to stem-controlled vowel harmony, see Baković 2000, 2001.) First, harmony is never systematically determined by a morphological unit other than the root/stem; the claim that stem control is due to cyclicity explains this fact. Second, it is explained why vowel harmony is only (apparently) directional in a language when that language has only suffixes or only prefixes: harmony seems to come from the left when there are only suffixes, because the root/stem is always on the left; harmony seems to come from the right when there are only prefixes, because the root/stem is always on the right.

2. Harmonic Pairing

The above descriptions of stem-controlled and dominant-recessive systems are generally accurate under what one might call ‘ideal’ conditions where at the very least each of the vowels in the word is *harmonically paired*, as defined in (4).

- (4) A vowel x in (the vowel inventory of) a language L with a harmonic feature $[\pm hf]$ is *harmonically paired* iff there is another vowel in L that differs from x only in terms of $[\pm hf]$.

For example, in the languages in (1) and (2), all high and mid vowels are harmonically paired but the low vowel is not, because the low vowel is [−ATR] and there is no [+ATR] low vowel.³ This situation is depicted graphically in (5); arrows between vowels indicate a harmonic pairing relationship.

³ Akan has such a vowel in (gradient) postlexical contexts (Clements 1981, Kiparsky 1985).

(5) Harmonic pairings in Akan, Tangale, Maasai, and Turkana

	[-back]		[+back]		
high vowels	i	ɨ	ɯ	u	[+ATR]
	ɪ	ɨ	ʊ	ɤ	[-ATR]
mid vowels	e	ɛ	ɔ	o	[+ATR]
	ɛ	ɛ	ɔ	ɤ	[-ATR]
low vowel			a		[-ATR]

When there is a harmonically unpaired vowel in a word, it can exhibit one of a small number of properties due to the application of vowel harmony. One property is what is known as *opacity*. An opaque vowel is a harmonically unpaired vowel that blocks the propagation of vowel harmony, whether harmony is instigated by a root/stem vowel or by a (dominant) affix vowel. If the harmony-instigating vowel and an opaque vowel have different values of the harmonic feature, then they surface with those different values, resulting in a (predictably) disharmonic form as defined precisely in (6) below.

- (6) Let x be any vowel and y be the vowel in any word instigating $[\pm hf]$ harmony. x is *opaque* iff (a) x is not harmonically paired, (b) a harmonically paired vowel between x and y agrees with y in terms of $[\pm hf]$, and (c) a harmonically paired vowel on the side of x opposite y agrees with x .⁴

Examples of opacity from each of the four languages in (1) and (2) are given in (7). In all four languages, the phonemic low vowel /a/ is opaque; specifically, this [-ATR] vowel blocks the propagation of [+ATR].

(7) Examples of opacity (harmonically unpaired opaque vowel is *italicized*)

Akan	/O + √bɪsɪa + ɪ/ 3SG + ask + PAST	→ [obɪsɪaɪ] 'he asked (it)'
Tangale	/√peer + na + n + gO/ compel + -PRX.LOC + 1SG + PERF	→ [peernangɔ] 'compelled me'
Maasai	/ε + √ɪput + a + ri + ie/ 3SG + fill + MA + N + APPL	→ [ɛɪputariyie] 'it will get filled up'
Turkana	/a + √pɛg + aa + n + u/ GEN + deny + HAB + SG + NOM	→ [apɛgaanu] 'denial'

A second property that a harmonically unpaired vowel can exhibit due to the application of vowel harmony is *transparency*, which is just like opacity except that vowels on the side of the transparent vowel opposite the harmony-instigating vowel harmonize with the instigator, not the opaque vowel. None of the languages under discussion here exhibit transparency, and so I will not discuss it further.⁵

⁴ Note that this definition needn't presuppose that the opaque vowel (x) and the harmony instigator (y) have different values of the harmonic feature $[\pm hf]$. If they happen to have the same value in some word, then there is simply no disharmony to speak of.

⁵ See Baković & Wilson (2000) on the approach to transparency that I advocate.

Vowel Harmony and Cyclicity in Eastern Nilotic

A third property that a harmonically unpaired vowel can exhibit due to the application of vowel harmony is what I call *re-pairing*. A re-paired vowel is a harmonically unpaired vowel that, unlike an opaque or transparent vowel, does alternate as otherwise expected under harmony between a [+*hf*] vowel and a [−*hf*] vowel. However, since a re-paired vowel is harmonically unpaired, the two alternants of a re-paired vowel differ with respect to at least one feature besides [+*hf*].

- (8) Let *x* be an [*αhf*] vowel and *y* be the vowel in any word instigating [+*hf*] harmony. *x* is *re-paired* iff (a) *x* is not harmonically paired, (b) *x* alternates, in agreement with *y*, with a [−*αhf*] vowel *z*, and (c) there exists a feature [+*f*] such that (i) [+*f*] ≠ [+*hf*], (ii) *x* is [+*f*], and (iii) *z* is [−*f*].

Re-pairing alternations due to vowel harmony are not at all uncommon. In Diola Fogy (Sapir 1965), the [+low] vowel [a] alternates with the [−low] vowel [ə] under dominant-recessive [+ATR] harmony; in Yokuts (Newman 1944), the [−back] vowel [i] and the [+back] vowel [u] alternate under stem-controlled [+round] harmony; and in Turkish (Underhill 1976), the [+low] vowel [a] and the [−low] vowel [e] alternate under stem-controlled [+back] harmony.

Of particular interest here is the fact that Maasai and Turkana exhibit re-pairing in addition to opacity. The same harmonically unpaired low vowel /a/, which was shown to sometimes be opaque in (7), is also sometimes re-paired, alternating between a [+low, −ATR] vowel [a] and a [−low, +ATR] vowel [o].⁶ The re-pairing alternation is found in contexts distinct from the opacity contexts in (7); specifically, a suffixal /a/ is re-paired when a [+ATR] instigator of harmony is in the stem of suffixation of the /a/ (i.e., to the left of the /a/), whereas /a/ is opaque when all [+ATR] instigators are in suffixes outside the stem of suffixation of the /a/ (i.e., to the right of the /a/). Examples of this alternation are given in (9).

- (9) Examples of re-pairing

a.	Maasai	/ɪn + √mudɔŋ + a/ FEM.PL + noun + PL	→	[ɪmudɔŋo] 'kinship'
	Turkana	/ɛ + √pup + aa + n + a/ 3 + obey + HAB + SG + VOI	→	[ɛpupooŋo] 's/he is obedient'
b.	Maasai	/ɪn + √lɪpɔŋ + a/ FEM.PL + noun + PL	→	[ɪlɪpɔŋa] 'full-grown female'
	Turkana	/ɛ + √pɛg + aa + n + a/ 3 + argue + HAB + SG + VOI	→	[ɛpɛgaana] 's/he is argumentative'
c.	Maasai	/ɛ + √ɪput + a + rɪ + ɪe/ 3SG + fill + MA + N + APPL	→	[ɛɪputariyie] 'it will get filled up'
	Turkana	/a + √pɛg + aa + n + u/ GEN + deny + HAB + SG + NOM	→	[apɛgaanu] 'denial'

⁶ There is an additional difference in terms of [+round] between the alternants that I ignore here in the interests of clarity. Everything I say about [+low] applies to [+round] as well.

The examples in (9a) show that an /a/ in a suffix surfaces re-paired as [o] when preceded by a [+ATR] instigator in the root/stem. The minimally different examples in (9b) show that this surface mid vowel does in fact alternate with the low [-ATR] vowel [a] when it is *not* preceded by a [+ATR] instigator. Finally, the examples of opacity in (9c), repeated from (7), show that re-pairing only occurs when a suffix /a/ is preceded, and not when only followed, by a [+ATR] instigator.

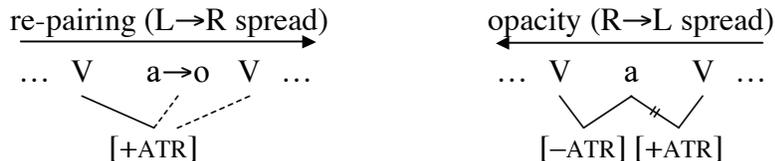
My focus in this paper is on the best explanation for the predictably asymmetrical behavior of /a/ in Maasai and Turkana just exemplified. I outline two approaches to this asymmetry, each making different claims as to its source: one in terms of directionality and the other in terms of cyclicity (as defined in (3) above). I then offer four arguments for the explanatory superiority of the cyclic approach.

3. Analytical Approaches

In this section I consider two basic analytical approaches to the question of what determines whether a suffixal /a/ is re-paired or opaque in Maasai and Turkana. One is a cyclic approach: whether /a/ is re-paired or opaque depends on whether or not a harmony instigator is *in the stem of suffixation of the /a/*. The other is a directional approach: whether /a/ is re-paired or opaque depends on whether or not a harmony instigator is *to the left of the /a/*.

Consider first the directional approach, advanced by (among others) Archangeli & Pulleyblank (1994) for Maasai and Albert (1995) for Turkana. Under this approach there are two [+ATR] harmony processes, one operating from left to right and the other from right to left. Unlike the left-to-right process, the right-to-left one is subject to some condition preventing its application to /a/.⁷ Thus, only the left-to-right process can cause re-pairing of /a/, while /a/ is opaque with respect to the right-to-left process. This is depicted graphically in (10).

(10) Re-pairing vs. opacity under the directional (spreading) approach



Under the cyclic approach, there is a difference in cyclicity between harmony and re-pairing. The harmony process itself is noncyclic, since it is dominant-recessive; a dominant [+ATR] suffix vowel can and does cause the vowel(s) of the stem to which it is suffixed to change (if necessary) to [+ATR]. However, whether a suffixal /a/ will be opaque or re-paired crucially depends on the cycle. The re-

⁷ This could be a condition on the (immediate) product of the process were it to apply to /a/ (e.g., a grounding condition against low [+ATR] vowels; Archangeli & Pulleyblank 1994:309) or one on the overall mapping from /a/ to [o] that would be necessary if the process were to apply to /a/ (a faithfulness constraint against changing the value of the feature [±low]; Albert 1995).

pairing process by which a suffixal /a/ becomes [o] — i.e., the change in [\pm low] in addition to the harmonic change in [\pm ATR] — only occurs under compulsion of harmony from the stem of suffixation, never under compulsion of harmony from a later suffix. So, if the /a/ is a part of a suffix attached to an already [+ATR] stem, then the /a/ is re-paired to [o]; if the /a/ is a part of a [-ATR] stem to which a suffix with an [+ATR] harmony instigator is suffixed, then the /a/ is opaque.

Empirically, the two approaches just outlined agree on the facts in (9). However, there are at least four arguments in favor of the cyclic approach. I discuss each of these arguments in turn in the next section.

4. Four Arguments for the Cyclic Approach

4.1. One: The Position of /a/ Relative to the Harmony Instigator

The first argument is that the cyclic approach predicts rather than stipulates the correct relationship between (i) suffixal /a/ being re-paired vs. opaque and (ii) the position of suffixal /a/ relative to a harmony instigator. Under the cyclic approach, the re-pairing process (the change in [\pm low]) is cyclic while the harmony process (the change in [\pm ATR]) is not. From this it follows that when a suffixal /a/ and its stem of suffixation disagree in [\pm ATR], the incompatibility will be resolved by effecting changes (those of harmony and re-pairing) on the suffixal /a/. However, a suffixal /a/ will not be re-paired when it only disagrees in [\pm ATR] with a following suffix; since /a/ is still harmonically unpaired, it will be opaque.

This does not follow directly from the directional approach. Under the directional approach, it is an arbitrary fact of the two directional harmony processes which one causes re-pairing of /a/ and which one doesn't. Consider the reverse situation, that the left-to-right process is subject to a condition preventing its application to /a/ whereas the right-to-left process is unencumbered. This would allow one to generate a pattern that is essentially the opposite of the pattern found in Maasai/Turkana — a pattern in which /a/ is re-paired where it is opaque in Maasai/Turkana and vice-versa. No such pattern seems to be attested.

There is thus an important typological consequence here: the pattern just described is predicted to be an impossible one under the cyclic approach. This is because re-pairing can be either cyclic or noncyclic. If it is cyclic, then we get the Maasai/Turkana pattern; if it is non-cyclic, then we get a pattern in which /a/ is always re-paired.⁸ The directional approach predicts these two patterns in addition to the unattested pattern described above. The way in which each of the three patterns is (or is not) predicted under the two approaches is summarized in (11).

⁸ This pattern corresponds to Diola Fogany (Sapir 1965), where low [a] alternates with mid [ə].

(11) Typological predictions of the two approaches

	<i>Pattern</i>	<i>Directional approach</i>	<i>Cyclic approach</i>
a.	Maasai/Turkana	right-to-left harmony cannot apply to /a/	re-pairing is cyclic
b.	/a/ always re-paired	both harmony processes can apply to /a/	re-pairing is noncyclic
c.	the opposite of Maasai/Turkana	left-to-right harmony cannot apply to /a/	impossible

A fourth pattern, one in which /a/ is always opaque, is predicted to exist by both approaches: under the directional approach, neither harmony process can apply to /a/; under the cyclic approach, one simply does not posit a re-pairing process in the first place. I do not at present know whether this prediction is attested.

In sum, the cyclic approach makes a desirably more restrictive claim than does the directional approach about the typology of possible dominant-recessive harmony patterns, and is therefore to be preferred.

4.2. Two: Interaction of Harmony and Re-pairing

The second argument for the cyclic approach has to do with the interaction between harmony and re-pairing. Under the cyclic approach, it is crucial that there be a process in addition to harmony that can be said to be cyclic, since harmony itself is dominant-recessive and therefore noncyclic. As noted above, /a/ is predicted to always be opaque if there is no re-pairing process. So, again, there is a restrictive typological consequence to the cyclic approach: a pattern just like that of Maasai/Turkana except that /a/ is not re-paired is predicted not to exist.

Under the directional approach, on the other hand, it is simply an accident that an /a/ which undergoes left-to-right harmony also undergoes re-pairing. It could just as simply have undergone harmony and surfaced as a low [+ATR] vowel. This is precisely because there are two separate harmony processes under the directional approach; the fact that both of them fail to create a low [+ATR] vowel, one via opacity and the other via re-pairing, is completely accidental.

4.3. Three: Two Directions of Harmony

The third argument in favor of the cyclic approach is the fact that it predicts rather than stipulates why harmony applies in both directions rather than in just one. The two ‘directions’ of harmony are only *apparent* under the cyclic approach, being distinguished by the cycle itself. Harmony processes can therefore be claimed to be inherently bidirectional, which Clements (1976 *et seq.*) originally argued is the right claim to make. Under the directional approach, there is no principled reason why Maasai and Turkana have two [\pm ATR] harmony processes in the first place, one applying in one direction and the other applying in the other direction. The

two processes are completely independent of one another, and so the prediction is that a grammar could have just one of these two independent processes, applying harmony in only one direction. Languages with unidirectional vowel harmony systems — that is, without any morphological motivation such as stem control coupled with a lack of prefixes — do not seem to exist.⁹

4.4. Four: Additional Evidence for Cyclicity in Turkana

So far I have offered what I consider to be three typological arguments for the cyclic approach: the set of possible patterns predicted by this approach are restricted in three ways that both explain crucial aspects of the facts at hand and appear to be typologically desirable. The fourth and final argument for the cyclic approach concerns an interesting additional set of facts from Turkana. These facts independently require that the re-pairing process be cyclic, as originally pointed out by Dimmendaal (1983:23ff; see also Albert 1995) and thus the burden of proof rests on the directional approach: since cyclicity is independently necessary, there appears to be no need for directional harmony processes.¹⁰

Even though [+ATR] vowels are the usual harmony instigators of Turkana, there is a small set of suffixes with a [–ATR] vowel such that when one of them is suffixed to a stem with an otherwise dominant [+ATR] vowel, the suffix vowel causes the [+ATR] vowel to become [–ATR]. The “anti-dominant” behavior of one of these suffixes is shown in (12) below; note that even the usually dominant [+ATR] vowels of the root ‘give birth’ surface as [–ATR].¹¹

(12) “Anti-dominant” [–ATR] harmony in Turkana

$/a + k + \sqrt{\text{ido}} + \text{un} + \underline{\text{et}}/$	→	$[\underline{\text{akidounet}}]$
GEN + K + give birth + VEN + INST-LOC		‘birth’

Of immediate interest is the interaction between one of these anti-dominant [–ATR] suffix vowels and a dominant [+ATR] vowel when there is an /a/ between them. Relevant examples are given in (13). The first two examples establish that the suffix vowel glossed as ‘E’ is an /a/ that (expectedly) surfaces re-paired as [o] when suffixed to a root like ‘drop’ with dominant [+ATR] vowels.¹² The example in (13b) is also the stem of suffixation for the anti-dominant suffix glossed as

⁹ I am not well acquainted with harmony processes limited to morphophonological domains smaller than the word, such as height harmony in some Bantu languages. If these show some evidence of directionality independent of the cycle, one would want to examine these domains to see if the mechanisms responsible for their existence might help to explain the apparent directionality.

¹⁰ Levergood (1984) argues for the cycle in her analysis of Maasai vowel harmony, but not with respect to re-pairing; see Levergood’s work as well as Baković 2000 (esp. pp. 232-236) for details.

¹¹ Double underlining indicates the [–ATR] suffix vowel and its effect on the rest of the word.

¹² The gloss ‘E’ stands for ‘epipatetic vowel’. According to Dimmendaal (1983:203-204), this suffix is phonologically regular but serves no morphosyntactic function, though it probably once did (like the ‘moveable k’ prefix, glossed here as ‘K’).

‘VOI’ in (13c). Note what happens in this last example: all vowels are [−ATR], and the vowel of the ‘E’ suffix surfaces not as [a] but as [ɔ].

(13) Dominance, anti-dominance, and /a/ in Turkana

- | | | | |
|----|---|---|-----------------------------------|
| a. | $/a + k + \sqrt{\text{ipud}} + a + \text{kin}/$ | → | $[\text{akipudakɪn}]$ |
| | GEN + K + trample + E + DAT | | ‘to trample’ |
| b. | $/\varepsilon + \sqrt{\text{ibus}} + a + \text{kin}/$ | → | $[\text{eibusokɪn}]$ |
| | 3 + drop + E + DAT | | ‘it has fallen down’ |
| c. | $/\varepsilon + \sqrt{\text{ibus}} + a + \text{kin} + \underline{a}/$ | → | $[\underline{\text{eibusokina}}]$ |
| | 3 + drop + E + DAT + VOI | | ‘it has thrown itself down’ |

The analysis of these facts under the cyclic approach is straightforward. In the cyclic direction, [+ATR] harmony causes the /a/ of the ‘E’ suffix to become [+ATR] because it can be cyclically re-paired, as in (13). This form serves as the stem of suffixation for the ‘VOI’ suffix with the anti-dominant [−ATR] vowel in (13). The disagreement between the anti-dominant vowel and the vowels in the stem of suffixation is resolved in favor of the anti-dominant vowel, which is possible in Turkana because harmony is noncyclic. However, since re-pairing is cyclic, the vowel of the ‘E’ suffix cannot be changed back to [a]. [−ATR] harmony in the anti-cyclic ‘direction’ can thus *only* produce [ɔ] under these circumstances.

The correct result is also possible under the directional approach, but *if and only if* left-to-right harmony is *crucially ordered* before the process responsible for anti-dominant [−ATR] harmony. When left-to-right harmony applies, the low vowel of the ‘E’ suffix is re-paired, becoming [+ATR, −low]; then [−ATR] harmony applies, changing this vowel back to [−ATR] but not back to [+low]. The result is thus the correct vowel [ɔ]. The directional analysis thus has two related disadvantages when compared with the cyclic analysis. First, the facts in (13) must be stipulated (via extrinsic ordering) under the directional analysis, whereas they follow automatically from the cyclic analysis (via the intrinsic ordering imposed by the cycle). Second, the directional analysis presupposes a theory in which processes may be extrinsically ordered with respect to each other, whereas the cyclic analysis is also compatible with a theory that only countenances intrinsic ordering imposed by the cycle. The cyclic analysis is clearly to be preferred.

5. A Fly in the Ointment

There is an empirical difficulty with the cyclic approach that should be noted, however. The directional and cyclic approaches agree on the facts in (9) but differ in their predictions of the behavior of /a/ in prefixes. The directional approach predicts that an /a/ in a prefix will be re-paired only if a harmony instigator is somewhere to the left of it (i.e., in a preceding prefix). If the only harmony instigator(s) is/are somewhere to the right of it (in a following prefix, root, or suffix), it will be opaque. The cyclic approach, on the other hand, makes essen-

Vowel Harmony and Cyclicity in Eastern Nilotic

tially the opposite prediction. An /a/ in a prefix will be re-paired if a harmony instigator is in the stem of prefixation; i.e., in a following prefix or root (or even in a suffix, if suffixed hierarchically ‘inside’ the prefix). If the only harmony instigator(s) is/are in a preceding prefix (or in a suffix, if suffixed hierarchically ‘outside’ the prefix), the /a/ is predicted to be opaque.

It seems that the directional approach is consistent with the full set of facts in these two languages, while the cyclic approach would need to be augmented somehow. Specifically, a prefixal /a/ is opaque when a harmony instigator is anywhere to the right of it, as shown by the data in (14) below.

(14) Opacity in prefixes

Maasai	a.	<i>/a + √rɔk + u/</i>	→	[aroku]
		1SG + black + INCEP		‘I become black’
	b.	<i>/a + √duŋ + akɪn + ie/</i>	→	[aduŋokinie]
		1SG + cut + DAT + APPL		‘s/he will hide him/herself’
Turkana	c.	<i>/a + √lilim + u/</i>	→	[alilimu]
		GEN + cold + NOM		‘coldness’
	d.	<i>/a + √tur + aan + a/</i>	→	[aturoonu]
		GEN + agile + HAB + NOM		‘agility’

There is an important systematic gap in the facts to consider, however: there are no dominant prefix vowels in either of these languages; that is, no prefix vowel is ever a harmony instigator. So, what happens to a prefixal /a/ when there is a harmony instigator preceding it, or one in a following prefix, cannot be tested. Given only the facts in (14), then, the problem with the cyclic approach amounts to the following: /a/ is never re-paired in prefixes, even though it is expected to be based on the behavior of /a/ in suffixes. Therefore, the solution to this problem is to somehow limit the re-pairing process to apply only in suffixes. Such a solution would ideally be derivable from independently necessary principles. Maasai and Turkana are not alone in having no dominant prefix vowels; it appears that *no* language with dominant-recessive harmony has dominant prefix vowels. While I do not know of (and do not offer) a principled account of this apparent universal, the unexpected behavior of prefixes more generally is clearly a vital area of further research into the problems posed by and related to the facts in (14).

6. Conclusion

I hope to have convinced the reader that there are more and stronger arguments for a cyclic as opposed to a directional approach to the asymmetrical behavior of /a/ in Maasai and Turkana. Since three of the four arguments presented rest on the restrictive typological claims made by the cyclic approach, it remains to be seen whether future research in this area will confirm or refute those arguments.

References

- Albert, Chris. 1995. Derived environments in optimality theory: A case study in Turkana. Ms., University of California, Santa Cruz.
- Archangeli, Diana, and Douglas Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.
- Baković, Eric. 2000. Harmony, dominance, and control. Ph.D. diss., Rutgers University, New Brunswick, NJ. [ROA-360, Rutgers Optimality Archive, <http://ruccs.rutgers.edu/roa.html>.]
- Baković, Eric. 2001. Arguments for a cyclic approach to vowel harmony. Ms., University of California, San Diego.
- Baković, Eric, and Colin Wilson. 2000. Transparency, strict locality, and targeted constraints. In *Proceedings of WCCFL 19*. [ROA-430, Rutgers Optimality Archive, <http://ruccs.rutgers.edu/roa.html>.]
- Clements, George N. 1976. Vowel harmony in nonlinear generative phonology: An autosegmental model. [Reproduced by IULC, 1980.]
- Clements, George N. 1981. Akan vowel harmony: A nonlinear analysis. *Harvard Studies in Phonology, Vol. II*: 108-177.
- Dimmendaal, Gerrit Jan. 1983. *The Turkana Language*. Dordrecht: Foris.
- Kidda, Mairo E. 1985. *Tangale Phonology: A Descriptive Analysis*. Berlin: Dietrich Reimer Verlag.
- Kiparsky, Paul. 1985. Some consequences of lexical phonology. *Phonology Yearbook 2*: 83-138.
- Levergood, Barbara. 1984. Rule governed vowel harmony and the strict cycle. *Texas Linguistic Forum 24*: 33-55 / *NELS 14*: 275-293.
- Newman, Stanley. 1944. *Yokuts Language of California*. Viking Fund Publications in Anthropology, No. 2, New York.
- Sapir, J. David. 1965. *A Grammar of Diola Fogny*. West African Language Monographs 3, Cambridge University Press, Cambridge.
- Schachter, Paul, and Victoria Fromkin. 1968. A phonology of Akan: Akuapem, Asante, and Fante. Working Papers in Phonetics 9, Department of Linguistics, University of California, Los Angeles.
- Tucker, Archibald N., and Tompo ole Mpaayei. 1955. *A Maasai Grammar, with Vocabulary*. London: Longmans, Green and Co.
- Underhill, Robert. 1976. *Turkish Grammar*. Cambridge, MA: MIT Press.

Department of Linguistics
University of California, San Diego
9500 Gilman Drive, #0108
La Jolla, CA 92093-0108

bakovic@ling.ucsd.edu

Constructional Effects of *Just Because ... Doesn't Mean ...*

EMILY M. BENDER and ANDREAS KATHOL
University of California, Berkeley

0. Introduction

This paper is concerned with the construction illustrated in (1), which we will refer to as the “JB-X DM-Y construction”.

- (1) Just because we live in Berkeley doesn't mean we're left wing radicals.

We will argue that this construction combines semantic and syntactic quirks that necessitate a constructional analysis. Further, we will show that specifying the pragmatic properties of the construction (and in particular the presuppositions that introduces) allows for a particularly elegant account of the construction's distribution. This, in turn, provides further support for the sign-based view of grammar in which syntactic constraints interact on an equal footing with semantic and pragmatic information.

1. Semantics of *Just Because* Sentences

As is well-known (cf. Jespersen 1949:399), *because*-clauses in English are in principle ambiguous between a causal and an inferential reading, as illustrated in the contrasting pair of sentences in (2):

- (2) a. The ground is wet **because**_{causal} it has rained.
b. It has rained (= must have rained) **because**_{inferential} the ground is wet.

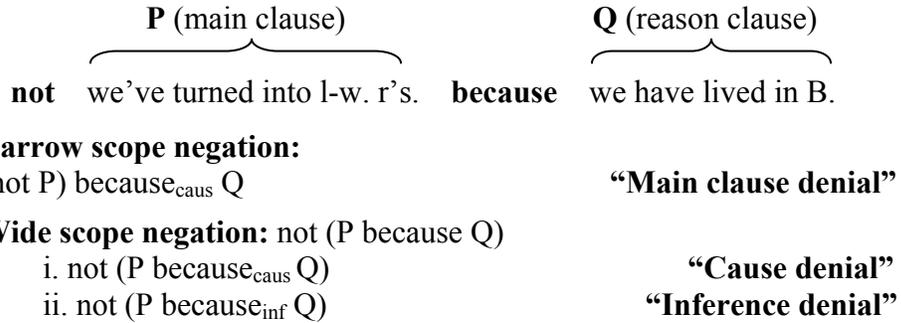
In (2a) the rain is understood as causally responsible for the wetness of the ground. In (2b), the wet ground is taken to license the abductive inference that there presumably has been rain that caused the wet ground to come about. Following Hirose 1991 we will refer to the two construals as the *causal* and *inferential* readings of *because*-clauses, respectively.

On the causal interpretation, reason clauses introduced by *just because* denote sufficient reasons for why a certain state of affairs holds. Thus, in (3), living in Berkeley is understood as causally responsible for becoming left-wing radicals, either potentially among others (3a) or as the single sufficient reason (3b):

- (3) a. We've turned into left-wing radicals **because** we have lived in Berkeley (for a year).
 b. We've turned into left-wing radicals **just because** we have lived in Berkeley (for a year).

When the main clause of such sentences occurs negated, as in (4), an ambiguity arises. This ambiguity is based on the scope of the negation on the one hand and the cause/inference ambiguity of *because* on the other. If the negation takes narrow scope, only the main clause is negated and only the causal reading appears possible. This is shown in (4a). When the negation takes wide scope, the ambiguity seen earlier with *because* gives rise to two readings. First, the causal connection may be denied, for instance if an alternative cause for the main clause is assumed. For example, something other than our residing in Berkeley caused us to become left-wing radicals. We will call this the “cause denial” as in (4bi). Alternatively, the existence of an inferential connection between reason and main clause may be negated, as shown in (4bii). This reading, which we call “inference denial” also strongly implicates that the main clause does not hold in the first place. These various readings are usually disambiguated via intonation.

- (4) We **haven't** turned into left-wing radicals **because** we have lived in Berkeley for a year.



The same ambiguity seems to also exist for *just because* reason clauses illustrated in (5), again notwithstanding intonational differences.

- (5) We **haven't** turned into left-wing radicals(,) **just because** we have lived in Berkeley for a year.
- | | | |
|---|--|---------------------------|
| a. Narrow scope negation: | | |
| (not P) just because _{caus} Q | | Main clause denial |
| b. Wide scope negation: not (P just because Q) | | |
| i. not (P just-because _{caus} Q) | | Cause denial |
| ii. not (P just-because _{inf} Q) | | Inference denial |

The order between reason and negated main clause may also be reversed. With simple *because* clauses, this order only allows a narrow scope reading of the negation. No wide scope reading under either the causal or the inferential reading of *because* appears to be possible, as illustrated in (6):

- (6) **Because** we have lived in Berkeley for a year we **haven't** turned into left-wing radicals.
- | | |
|--|--|
| $\underbrace{\hspace{10em}}$
Q (reason clause) | $\underbrace{\hspace{10em}}$
P (main clause) |
| because we have lived in B. | not we have turned into l-w. r's. |
- a. **Narrow scope negation:** (not P) $\text{because}_{\text{caus}}$ Q
 b. **No wide scope negation**

If an adverbial clause in initial position is structurally higher than the main clause, then the difference in negation scope for the different orders is predicted.

In contrast, preposed *just because* clauses continue to allow for both a narrow and a wide scope construal of the negation. However, now the wide scope negation only allows for the inference denial interpretation.

- (7) **Just because** we have lived in Berkeley for a year we **haven't** turned into leftwing radicals.
- | | |
|--|---------------------------|
| a. Narrow scope negation: (not P) $\text{just-because}_{\text{caus}}$ Q | Main clause denial |
| b. Wide scope negation: not (P just-because Q) | |
| i. No cause denial | |
| i.e., not available: not (P $\text{just-because}_{\text{caus}}$ Q) | |
| ii. not (P $\text{just-because}_{\text{inf}}$ Q) | Inference denial |

A wide scope reading of the negation with cause denial construal no longer seems available. That is, by saying (7), a speaker either asserts (main clause denial) or strongly implicates (inference denial) that he/she is not a left-wing radical. With the main clause denial reading, living in Berkeley is claimed to be sufficient for this to come about, whereas in the inference denial reading, the speakers reject the idea that their residence should license conclusions about their political opinions.

Given the fact that initial simple *because* clauses do not allow wide scope negation (cf. McCawley 1988), one may expect initial *just because* clauses to behave similarly. The fact that the latter do allow for wide scope negation is therefore unexpected—the ‘ordinary’ mechanisms of grammar do not provide for this reading. In order to account for the reading, we posit a construction (in particular, a specialized subtype of head-modifier constructions) which calls for a *just because* adjunct preceding a negated main clause, and specifies that the negation in the main clause should take scope over the adjunct.

Examples such as (1) are licensed by a further subtype of this construction, as discussed in the next section.

2. Lexicalizing Inference Denial

The inference denial reading of sentences of the form (8) can be broken down into the components in (9):

- (8) Just because Q, not P.
- (9) • P cannot be inferred from Q.
 • not P (implicature)

Furthermore, sentences of the form in (8) carry at least two presuppositions: that Q holds,¹ and that someone (by default the addressee) believes that P can be inferred from Q.

In the subclass of *just because* constructions which we will focus on in this paper, the first component of the meaning (that P cannot be inferred from Q) is lexicalized in a negated verb such as *mean*. In such sentences, only the inference denial reading appears to be possible.² As is illustrated in (10), the choice of main clause subject in such cases is rather restricted. Demonstrative *that*, understood as referring to the propositional core of the *just because* clause, appears best, followed by *it*. Other choices seem relatively degraded, as is illustrated in (10c,d):

- (10) a. Just because we live in Berkeley **that** doesn't mean that we're left-wing radicals.
 b. Just because we live in Berkeley, **it** doesn't mean that we're left-wing radicals.
 c. ?Just because we live in Berkeley, **this** doesn't mean that we're left-wing radicals.
 d. ?Just because we live in Berkeley, **that fact** doesn't mean that we're leftwing radicals.

Another possibility involves simply juxtaposing the *just because* clause and the *doesn't mean* VP, as seen earlier in (1), repeated below:

¹ As Hirose 1991:31 points out, this presupposition seems to be a general property of preposed *because* clauses.

² The main clause denial reading is possible in superficially similar examples where the demonstrative subject, which must be overt in these cases, refers to a proposition other than that expressed by JB-X:

- (i) a. [We inherited \$500,000].
 b. [Just because we live in Berkeley], that doesn't mean that we can afford a nice house.
 'Living in Berkeley is sufficient reason for the idea that inheriting \$500,000 does not imply being able to afford a nice house.'

- (11) **JB-X** **DM-Y**
{ **Just because** we live in Berkeley { **doesn't mean** we're left-wing radicals. }

It is natural to think of such expressions as one further step in the grammaticalization of the inference denial interpretation and therefore as being licensed by a particular subconstruction of a more general inference denial construction. Examples of this type will constitute the focus of the remainder of this paper, and we will refer to them as JB-X DM-Y sentences.

We have noted above that the preservation of the inference denial reading despite the preposing of JB-X argues for a constructional analysis of JB-X DM-Y. On this analysis, a particular construction licenses this pairing of form and meaning which is not predicted by the rest of the grammar of English. The cases discussed in this section above would seem to call for a more specialized subconstruction. This small hierarchy of constructions can be conceptualized as in (12):

- (12)
- ```

graph TD
 A[head-adj-ph] --- B[JB-inference-denial]
 A --- C[...]
 B --- D[JB-X-not-Y]
 B --- E[JB-X-DM-Y]

```

*JB-inference-denial* is a subtype of *head-adj-ph* and it encodes what is common to both subtypes: the preposing of the *just because*-clause and the inference denial semantics. *JB-X-not-Y* need not add any further constraints. In particular the *just because* clause acts as a modifier that combines with a regular clause that does not contain a predicate of inference. It contrasts with *JB-X-DM-Y* which licenses sentences with a full main clause part like (10) (which we will call “clausal JB-X DM-Y”) and those in which *just because* is juxtaposed with a surface VP, as in (11) (which we will “predicate JB-X DM-Y”). It may seem surprising at first for predicate JB-X DM-Y sentences to be licensed by an eventual subtype of *head-adj-ph*. However, as we briefly discuss below, we believe that JB-X retains its modifier status even in these cases.

### 3. Constructional Properties of JB-X DM-Y

The discussion of JB-X DM-Y sentences in the literature (specifically Hirose 1991 and Holmes and Hudson 2000) either implicitly or explicitly assumes that there are (at least) three properties that need to be specified in the description of this construction. First, in the case of predicate JB-X DM-Y constructions, it is the JB-X part that constitutes the subject. Second, the predicate in the DM-Y part has to occur negated. Third, the only type of predicate that can head the DM-Y part is *mean*, or at least a very small set of predicates. We find that a closer examination of the data provides evidence against all three of these assumptions.

### 3.1. Subject of DM-Y

Both Hirose 1991 and Holmes and Hudson 2000 explicitly adopt the idea that in predicate JB-X DM-Y cases like (11), the *just because* clause itself constitutes the subject of the following predicate. Such an analysis entails that JB-X DM-Y constructions have to be considered a syntactically heterogeneous class. If the main clause contains a pronominal subject, the *just because* clause is an adjunct, otherwise it is a subject.

An alternative possibility is for the *just because* clause to always be an adjunct. In predicate JB-X DM-Y cases, the construal of the *just because* clause as the subject of *mean* is not the result of an ordinary subject-predicate structure, but instead is mediated by some other (construction-specific) means. We believe that this issue is in principle subject to empirical study by comparing the behavior of *just because* clauses to other clausal subjects in a number of environments that are reserved for subjects. While native speaker judgments in this area are notoriously difficult to evaluate, we have presented preliminary experimental evidence elsewhere (Bender and Kathol 2001) that indeed argues against the subject status of *just because* clauses. If further study confirms these results, our proposed analysis would constitute independent evidence for the idea of constructionally licensed unexpressed subjects in English finite clauses, as recently proposed for a subtype of tag questions by Kay (2000).

### 3.2. Negated Predicate in DM-Y

At first glance, it would seem that the *JB-X-DM-Y* construction should also specify that *mean* in the head daughter be negated. Thus, it is hard to imagine a context that would make an example without negation, such as (13), sound acceptable:

(13) \*Just because we live in Berkeley means we're left-wing radicals.

However, on closer examination it turns out that the lexicalization of inference denial does not require explicit negation of the *mean* predicate in the form of *doesn't mean*. Consider first the following corpus examples in which the negation takes another form:<sup>3</sup>

(14) @Yet, just because some people cannot distinguish between serious and hypothetical risks **hardly means** that knowledgeable Republicans cannot muster the courage to speak out for health.

(15) @ "Just because someone has a black belt **means nothing**," said Jones.

---

<sup>3</sup> The symbol @ before an example sentence indicates that it is an attested example. All such examples here are from the North American News Text Corpus, available from the Linguistic Data Consortium: <http://www ldc.upenn.edu>

- (16) @“You **haven't said** - and I'm not saying - that just because a person makes that kind of money **means** there is waste, fraud and abuse,” Bilirakis said.

In fact, JB-X DM-Y sentences appear to have roughly the same distribution as negative polarity items (NPIs): They are licensed in polar questions (17),<sup>4</sup> antecedents of conditionals (18), and complements of implicit negative predicates (19).

- (17) @“Just because a guy has bleached hair, winter tan, speaks slowly and is pleasant to the point of being vacuous,” asks a pointed essay in the magazine, “**does that mean** he's a surfer?”
- (18) **If** just because we live in Berkeley **means** we're left-wing radicals, you have some serious misconceptions about our city.
- (19) I **doubt** that just because they live in Berkeley **means** they're left-wing radicals.

Like NPIs, the negation for JB-X DM-Y sentences can be supplied by sentence initial *like*, which functions to express irony and hence indirectly negates the contents of what follows.<sup>5</sup>

- (20) a. **Like** just because we live in Berkeley **means** we're left-wing radicals!  
b. Bill Gates received a huge tax return this year. **Like** he needs any more money!

However, on closer inspection, the parallelism between JB-X DM-Y sentences and NPIs breaks down. First, if there is no lexical indicator of irony and the negation of the literal content is entirely a pragmatic effect (possibly aided by intonation), regular NPIs are no longer licensed, as shown in (21a). In contrast, JB-X DM-Y still appears to be possible, as illustrated in (21b):

---

<sup>4</sup> Note that in polar questions, the subject of the *mean* predicate must be overt, thus the following is impossible:

- (i) \*... **does mean** he's a surfer?

It may be thought that a subject-less approach to predicate JB-X DM-Y of the kind briefly discussed in section 4.1 falsely predicts (i) to be grammatical. However, this is not so if subject-auxiliary (SAI) constructions are generally required to contain a phonologically expressed subject. See also Fillmore 1999 on SAI constructions.

<sup>5</sup> Thanks to Chuck Fillmore for this particular example and to Michael Israel and Paul Kay for general discussion of JB-X DM-Y and NPI-licensing.

- (21) a. (So, let me get this straight, )  
just because we live in Berkeley **means** we're left-wing radicals.  
b. (So, let me get this straight, )  
\*he needs any more money.

More tellingly, JB-X DM-Y sentences appear to be licensed by any environment that distances the speaker from the belief that X in fact implies Y.<sup>6</sup>

- (22) a. Kim **seems to believe** that just because we live in Berkeley **means** we're left wing radicals.  
b. \*Kim **knows** that just because we live in Berkeley **means** we're left wing radicals.

The proper generalization behind the above examples appears to be that the *JB-X-DM-Y* construction contributes the information that the speaker believes that Y cannot be inferred from X. This contribution interacts with the lexical content of the sentences and the way in which they are used to license the pattern of judgments discussed above:

In sentences like (11), 'Y can't be inferred from X' is directly encoded by the lexical expressions (*doesn't mean*). Furthermore, this is understood to be consistent with the speaker's beliefs, since the speaker is asserting it. In sentences like (13), the surface string expresses 'Y can be inferred from X' and, since the speaker asserts this, this must be what the speaker believes. The resulting conflict between this assertion and inference denial effect of the JB-X DM-Y construction as a whole makes such sentences infelicitous. On the other hand, in sentences like (21a), the surface string expresses 'Y can be inferred from X', but this is negated by the sarcastic use. The sarcasm thus indicates that the speaker believes that Y can't be inferred from X, and *JB-X-DM-Y* is felicitous. In sentences like (22), the speaker is attributing the belief that 'Y can be inferred from X' to Kim. By using *JB-X-DM-Y* to express this information, the speaker is also conveying that s/he believes that Y cannot be inferred from X. Note that when the matrix verb is changed to a factive verb like *know*, the sentence becomes unacceptable. Interestingly, the exact opposite behavior arises if the embedded clause is negated, as in (23). Here, the possibilities for the matrix verb are the mirror image of what they were in (22). *Seems to believe* distances the speaker from the content of the complement of *believe*. Since this would mean that the speaker believes 'Y can be inferred from X', this use of *JB-X-DM-Y* is infelicitous. In contrast, factive *know* is fine here, as shown in (23b):

---

<sup>6</sup> Thanks to Abby Wright for pointing out this type of example to us.

- (23) a. \*Kim seems to believe that just because we live in Berkeley **doesn't mean** we're left wing radicals.  
b. Kim knows that just because we live in Berkeley **doesn't mean** we're left wing radicals.

Finally, this analysis predicts that the polar question examples like (17) above should have the flavor of a rhetorical question, that is, a question in which the speaker already knows the negative answer. We believe that this is indeed the case and that JB-X DM-Y cannot be used if the speaker intends for the polar question to resolve a genuine issue.

Thus the apparent need for negation is actually due to a semantic/pragmatic contribution of the construction. However, this contribution interacts with the rest of the meaning of JB-X DM-Y sentences in what strikes us as unusual ways. In the most common case (sentences such as (11)), the constructional contribution ('The speaker believes that Y can't be inferred from X') appears redundant because this is exactly the meaning one would get from the meaning of the words and the way they are used. In other cases (such as (22)), the construction appears to be providing information beyond what is expressed in the words. In still other cases (such as (13)), the construction appears to be infelicitous because the construction contribution is incompatible with other aspects of the utterance meaning.

It is unclear to us at the moment exactly what kind of meaning this constructional contribution is. It is unlike presuppositions in that it is not backgrounded but rather asserted. It is unlike conversational implicatures in that it does not appear to be defeasible:

- (24) #Kim seems to believe that just because we live in Berkeley means we're left wing radicals, and I think I might just think so, too.

It may be a type of conventional implicature, if there exist conventional implicatures that are not backgrounded like presuppositions (cf. Karttunen and Peters 1979).

### **3.3. Lexical Variability/Constructional Stability**

Previous work on the JB-X DM-Y construction has either described the construction in terms of the selectional properties of a specific lexical element, i.e., *mean* (Holmes and Hudson 2000), or has allowed for very limited degree of lexical variation. Thus, Hirose (1991:18–19) mentions that in addition to inference predicates such as *mean* and *is no reason*, one can also find examples with *doesn't make*.

An initial informal survey of corpus examples drawn from North American News Text Corpus has revealed that the focus on *doesn't mean* is to some extent justified by the sheer numerical predominance of this item (about 85% of the surveyed subcorpus). Prototypical constructions of this kind occur about 14 times

more often than the second most frequent predicate (*doesn't make* with about 6% of occurrences).

At the same time, however, the degree of lexical variation is far greater than Hirose's discussion would lead one to expect. It also appears that by and large the type of predicates admitted into this construction is roughly the same as the range of meanings of either *mean* or *make*.

### 3.3.1. Variation on *Mean*

The range of predicates that appear to be related to senses of *mean* fall into three broad classes: predicates of inference (25)–(28), predicates of evidence (29)–(32), and predicates of (moral) justification (33)–(35). Notice that some of these predicates take non-clausal complements.

#### Predicates of inference

- (25) @“There are some issues that need to be resolved,” Mr. Blumenthal said, “but just because there is an investigation **by no means should be taken to infer** that any wrongdoing has occurred.”
- (26) @Just because a guy knocks out a hamburger in the first round **doesn't establish the fact** he's back.
- (27) @Ito said that just because the source had access to the less advanced tests **did not prove** that the source had access to the sock.
- (28) @So just because you meet with the “rep” in the cafeteria, union office or faculty room **doesn't imply** that your employer endorses the investments.

#### Predicates of evidence

- (29) @ “Just because a person has very high grades and looks like a model citizen **does not** always **indicate** that they are a fine human being,” he said.
- (30) @ “Just because other areas are doing okay, **is not a sign that** we in New England are doing badly,” said Gaal.
- (31) @ “Just because he's adopting a Republican agenda in a timely fashion **doesn't reflect** growing in the job,” said Gary Koops, deputy campaign director for Clinton's Republican challenger, Bob Dole.
- (32) @Just because there is profanity in a book **doesn't say** you condone or endorse that.

### **Predicates of justification**

- (33) @They emphasize that culture can and often must supersede instinct: that just because apes commit rape **in no way justifies** similar behavior in humans.
- (34) @Just because an officer sees a bulge **doesn't give him the right** to grab a student and search that student.
- (35) @“Just because we did a lousy job in fee-for-service **is not an excuse** to do a lousy job with HMOs,” Ms. Dallek said.

### **3.3.2. Variation on *Make***

In contrast to *mean* and its various related replacements, which focus on the way that a cognitive agent may establish an inference relation between two states of affairs, the sentences containing (*doesn't*) *make* emphasize a different kind of connection between the two states of affairs. Two examples from the corpus are given in (36) and (37):

- (36) @“Just because the doctor can't find out what's wrong with me **doesn't make** my back hurt any less,” Dr. Reed said.
- (37) @Just because McCamant or any analyst says a company is ripe to be acquired **doesn't make** it true.

In these examples, as well as the variations that follow below, the relation in question is more closely connected to a notion of causation according to conventions of society or natural law.

- (38) @I mean, just because we beat Phoenix **doesn't move** us into the Top 25 of the AP poll.
- (39) @Seifert said Monday that just because the doctor stamped Young's ticket **doesn't** necessarily **admit** him to the dance.
- (40) @“Just because the driver was a different race **does not qualify** it as a hate crime,” Pigott said.
- (41) @Just because Rosenthal was able to cope with reality on the job and acted normal in a video taken two days before the murder with his four-month-old daughter **does not mitigate** the diagnosis, Whaley said.
- (42) @Just because it has some setbacks and challenges this year **doesn't affect** that at all.

- (43) @Just because employees dislike each other **is not an automatic cause** for alarm.

Interestingly, in the following examples, the predicate (*preclude*, *negate*) is normally used to express the *lack* of a relation between two entities. Thus, the JB-X DM-Y construction is used to convey that contrary to conventional wisdom, a relevant connection does exist.

- (44) @Just because some land deal is being made **does not negate** the need for affordable housing in San Francisco.
- (45) @Just because I'm 65 **doesn't stop** me setting the target.
- (46) @Just because someone is involved in civic affairs and supports candidates **should not** automatically **exclude** them from conducting a business.

### 3.3.3. Residual Cases

Finally, in the following, we list additional examples, which do not seem to be related to any sense of *mean* or *make* in an obvious way.

- (47) @“And just because a place is a party school **is not a bad thing**,” Custard said.
- (48) @Just because the data scavengers have scraped it together and started to sell it doesn't **begin to answer** the question whether they own it—or whether it's right.
- (49) @Just because your parents are in the business **is not enough**, unless you have the desire.
- (50) @Just because the recogniser has little confidence in a particular character **need bear no resemblance** to whether or not that is the incorrect character in a misspelled word.

## 4. Conclusion

Makkai (1972:57) distinguishes two kinds of idioms:

IDIOMS OF ENCODING: “[Constructions] whose existence is justified by constant use by the majority of speakers ... [and which] compel the speaker to ENCODE in a certain way.”

IDIOMS OF DECODING: Constructions which “force the hearer to DECODE in a certain way”.

*JB-X-DM-Y* appears to have aspects of both. The constructional contribution to the meaning of *JB-X-DM-Y* sentences makes it an idiom of decoding. Makkai states that all idioms of decoding are also idioms of encoding, in that the special semantics is always attached to some form. In the case of *JB-X-DM-Y*, that form is somewhat underspecified. The construction stipulates the order of the two clauses, restricts the choice of subject for the second clause (to *it*, *that* or unexpressed), and restricts the choice of verbs in the second clause to some extent. The strong preference for *mean* in the second clause constitutes an overlaid idiom of encoding: that is, the knowledge that this is the way we usually say it.

## References

- Bender, Emily M., and Andreas Kathol. 2001. A constructional analysis of *just because...doesn't mean...* Talk presented at the First International Conference on Construction Grammar, April 6–8, 2001, UC Berkeley.
- Fillmore, Charles. 1999. Inversion and Constructional Inheritance. In Gert Welbuth and Jean-Pierre Koenig and Andreas Kathol (eds.) *Lexical and Constructional Aspects of Linguistic Explanation*. Stanford: CSLI Publications. 113–128.
- Hirose, Yukio. 1991. On a certain nominal use of *because*-clauses: Just because *because*-clauses can substitute for *that*-clauses does not mean that this is always possible. *English Linguistics*: 8. 16–33.
- Holmes, Jasper, and Richard Hudson. 2000. Just because X doesn't mean Y. Talk presented at the Linguistics Association of Great Britain, Spring Meeting, April 6–8, 2000, University College London.
- Jespersen, Otto. 1949. *A Modern English Grammar on Historical Principles. Part V*. London: George Allen & Unwin.
- Karttunen, Lauri, and Stanley Peters. 1979. Conventional implicature. In Choon-Kyu Oh and David A. Dineen (Eds.), *Syntax and Semantics 11: Presupposition*, New York: Academic Press.
- Kay, Paul. 2000. English Subjectless Tagged Sentences. Unpublished ms., UC Berkeley.
- Makkai, Adam. 1972. *Idiom Structure in English*. The Hague: Mouton.
- McCawley, James. 1988. *The Syntactic Constructions of English*. Chicago: University of Chicago Press.







*Modification of the Noun in Somali and in Cushitic*

- b. OK qálin-ka dhээр oo cusúb  
pen-det(m) big oo new  
'the big new pen' (two adjectives)
- (5) a. \* heesá-ha soomaaliy-éed cusúb  
songs-det(m) Somali-eed new
- b. OK heesá-ha soomaaliy-éed oo cusúb  
songs-det(m) Somali-eed oo new  
'the new Somali songs' (adjective+noun-*eed*)
- The coordination is necessary not only for modification by adjectives but also for modification by adjectives that are more complex, e.g. relative clauses are illustrated in (8) and (9).
- (6) a. \* kóob-ka [bulúug-ga ah] cusúb  
cup-det(m) [blue-det(m) is] new
- b. OK kóob-ka bulúug-ga ah oo cusúb  
cup-det(m) blue-det(m) is oo new  
'the new blue cup' (adjective+*ah*-relative)
- (7) a. \* gabdhá-haas [qurúx-da badán] [halkáas marayá ]  
girls-those beauty-det(f) much over there walk
- b. OK gabdhá-haas [qurúx-da badán] oo[halkáas marayá ]  
girls-those beauty-det(f) much oo over there walk  
'those very beautiful girls that are walking over there'  
(adj.noun + subj.relative)
- (8) a. \* wíil-ka yar [halkáas ku cayaarayá]  
boy-det(m) small there prep runs
- b. OK wíil-ka yar oo [halkáas ku cayaarayá]  
boy-det(m) small oo there prep runs  
'the little boy running there (far away)'  
(adjective+subj.relative)
- (9) a. \* wíil-ka yar [aan arkó]  
boy-det(m) small I see
- b. OK wíil-ka yar oo [aan arkó ]  
boy-det(m) small oo I see  
'the small boy that I see' (adjective+obj.relative)

- (10) qayb-tan yaab-kéeda [loo bogay] ee fiican  
 share-dem.f surprise-hers one+with be satisfied EE good  
 ‘this surprising and good way of sharing that one is satisfied with’  
 (in Morin (1986, p.114)) (relative clause + adjective)

The case of modification by a genitive and a relative is illustrated in (11c).

- (11) a. qálin-ka macállin-ka oo cusúb  
 pen-det(m) teacher-det(m) oo new  
 ‘the new pen of the teacher’ (genitive + adjective)
- b. kóob-ka macállin-ka oo [shaah-a ah]  
 cup-det(m) teacher-det(m) oo tea-det(m) is  
 ‘the tea cup of the teacher’ (genitive + ah-relative)
- c. waláal-ka Warsáme oo [halkáas kú cayaarayá]  
 brother-det(m) Warsame oo there prep. runs  
 ‘Warsame's brother who is running there (far away)’  
 (genitive + subj.relative)

The second property of noun modification in Somali is that the order of any two coordinated modifiers is free. Consider for instance modification by an adjective and an object relative. We see in (9b) that the order adj >> obj rel is good. Now as shown in (12b), the inverse order obj rel >> adj is also good.

- (12) a. qálin-ka cusúb oo macállin-ka  
 pen-det(m) new oo teacher-det(m)  
 ‘the new pen of the teacher’, cf. (11a) (adjective + genitive)
- b. wíil-ka [ aan arkó ] oo yar  
 boy-det(m) I see oo small  
 ‘the small boy that I see’, cf. (9b) (obj.relative + adjective)

Note that the modifier that appears first is perceived as more important or new.

- (13) a. qálin-ka macállin-ka oo cusúb  
 pen-det(m) teacher-det(m) oo new  
 OK: ‘the new pen of the TEACHER’ (genitive+adjective)
- b. qálin-ka cusúb oo macállin-ka  
 pen-det(m) new oo teacher-det(m)  
 OK: ‘the NEW pen of the teacher’ (adjective + genitive)



### 2.1. Afar

As expected for an SOV language, the word order inside the noun phrase in Afar is Modifier Noun. Various examples of modified nouns are given in (17). In all cases, the modifier precedes the head N.

|      |    |                               |  |                  |                                                              |
|------|----|-------------------------------|--|------------------|--------------------------------------------------------------|
| (17) |    | <i>Modifier</i>               |  | <i>Noun</i>      |                                                              |
|      | a. | woó<br>dem                    |  | 9ari<br>house    | ‘that house’<br><i>(demonstrative)</i>                       |
|      | b. | nabá<br>great                 |  | num<br>man       | ‘big/great/old man’<br><i>(adjective)</i>                    |
|      | c. | sagá-h<br>cow-gen             |  | iba<br>foot      | ‘a cow’s foot’<br><i>(genitive)</i>                          |
|      | d. | tidhdhigillé<br>it was broken |  | boddina<br>tooth | ‘tooth that was broken’<br><i>(relative)</i>                 |
|      |    |                               |  |                  | (exs a.b.d from Hayward 1998:625, ex c from Bliese 1977:277) |

The different types of modifiers follow a strict order: demonstrative >> adjective >> genitive (examples from Bliese 1977:25/285).<sup>4</sup>

|      |    |               |                  |                  |                                                  |
|------|----|---------------|------------------|------------------|--------------------------------------------------|
| (18) |    | <i>dem</i>    | <i>adj</i>       | <i>N</i>         |                                                  |
|      | OK | amá<br>this   | 9asá<br>red      | saró<br>clothing | sarté<br>I put on                                |
|      | *  | 9asá          | amá              | saró             | sarté<br><i>(* adj &gt;&gt; dem)</i>             |
| (19) |    | <i>dem</i>    | <i>adj</i>       |                  |                                                  |
|      | OK | wóo<br>that   | ni9ín<br>hot     | xán<br>milk      | mádhagin<br>don’t touch                          |
|      | *  | ni9ín         | wóo              | xán              | mádhagin<br><i>(* adj &gt;&gt; dem)</i>          |
| (20) |    | <i>adj</i>    | <i>gen</i>       | <i>N</i>         |                                                  |
|      |    | datá<br>black | awk-í<br>boy-gen | ko9só<br>ball    | ‘the boy’s black ball’<br><i>(adj.+genitive)</i> |

The examples in (18), (19), and (20), also show that multiple modification does not trigger coordination. These observations suggest that Afar has multiple slots for the different types of modifiers and that the modifiers occupy hierarchically ordered positions.

---

<sup>4</sup> Relatives are more complicated: rel >dem>N and dem>rel>N are both OK. (Bliese 1977:25,26)

*Modification of the Noun in Somali and in Cushitic*

**2.2. Harar Oromo**

In Harar Oromo, like in Somali, the word order inside the noun phrase is: Noun >> Modifier, as illustrated in (21) (examples from Owens 1985:87/ 103/ 107).

- |      |    |                      |                  |                                          |
|------|----|----------------------|------------------|------------------------------------------|
| (21) | a. | namá<br>men          | gaaríi<br>good   | ‘good men’<br>( <i>adjective</i> )       |
|      | b. | bif-níi<br>color-nom | sáree<br>dog/gen | ‘the dog’s color’<br>( <i>genitive</i> ) |
|      | c. | gaalá<br>camel       | xiyyán<br>my     | ‘my camel’<br>( <i>possessive</i> )      |

As in Afar, and unlike in Somali, however, the different types of modifiers appear in a fixed order: adjective >> genitive >> relative >> demonstrative, (Owens 1985:86).

- |      |    |                                  |            |                  |                  |
|------|----|----------------------------------|------------|------------------|------------------|
| (22) | a. | <i>N</i>                         | <i>adj</i> | <i>gen</i>       |                  |
|      |    | xeesuminná                       | gaaríi     | [ namiccá suní ] |                  |
|      |    | guests                           | good       | man that/gen     |                  |
|      |    | ‘that man’s good guests’         |            |                  | (Owens1985:104)  |
|      | b. | <i>N</i>                         | <i>adj</i> | <i>relative</i>  | <i>dem</i>       |
|      |    | makiináa                         | díimtúu    | [ at bítte ]     | sán              |
|      |    | car                              | red        | you buy2pst      | that             |
|      |    | ‘He saw that red car you bought’ |            |                  | (Owens 1985:132) |

Again as in Afar, multiple modification of the noun does not trigger the insertion of any coordinating particle in Harar Oromo.

**2.3. Iraqw**

The word order inside the noun Iraqw noun phrase is: Noun >> Modifier.

- |      |    |          |            |               |
|------|----|----------|------------|---------------|
| (23) | a. | <i>N</i> | <i>adj</i> |               |
|      |    | hhaysáa  | tléer      | ‘a long tail’ |
|      |    | tail     | long       |               |
|      | b. | garmoó   | úr         | ‘a big boy’   |
|      |    | boy      | big        |               |

We observe a strict ordering among different types of modifiers: genitive >> adjective (see (24) from Mous 1993:229), with relative clauses following all other modifiers (see (25) Mous 1993:231).

- (24) *N*                      *gen*                      *adj*  
 maká                      gadá                      ninakw                      ‘small forest animals’  
 animals(cs)              forests(cs)              small(p)
- (25) tsir’i tam ar              wák ar              aa              dakúus              i              káhh  
 bird(f) even indep.f.cs      one indep.f.cs 3perf      miss.3ms.past 3      be absent3fs  
 ‘A bird, even one, that he missed, does not exist’

Again, Iraqw allows multiple modification of the noun without recourse to a coordinating particle.

#### 2.4. Beja

In Beja the word order inside the noun phrase is variable: an indefinite adjective or a genitive precedes the head noun, (see (26)), while definite adjectives follow the head noun (see (27a)).

- (26) a.              win      kaam                      ‘a big camel’ (Almkvist:§90)  
                     big      camel                      (*adjective*)
- b.              masankoo-ti      biya                      ‘a harpstring’  
                                     harp-gen(f)      string                      (*genitive*)

This distribution implies in particular, that definite genitives and definite adjectives occupy different positions.

- (27) a.              *detN*                      *det adj*  
                     uu-kaam              uu-win                      ‘the big camel’ (Almkvist §90)  
                     det(m)-camel      det(m)-big                      (*definite noun+definite adjective*)
- b.              *det gen*                      *det N*  
                                     oo-taki                      oo-gaw                      ‘the house of the man’  
                                     det(m)-man      det(m)-house                      (*definite genitive+definite noun*)

In Beja, multiple modification does not trigger coordination.

- (28) uu-san-ii-ta                      meek      tuu-win-t  
                     det(m)-brother-gen-poss1s donkey      det-big-det(f)                      (Almkvist 1881:§119)  
                     ‘my brother’s big she-donkey’                      (*genitive + N + definite adjective*)

#### 2.5. Properties of noun modification in the languages examined

The properties of noun modification in the languages we examined are summarized in table (29).

*Modification of the Noun in Somali and in Cushitic*

|                                          |  |               |             |               |              |             |
|------------------------------------------|--|---------------|-------------|---------------|--------------|-------------|
| (29)                                     |  | <u>Somali</u> | <u>Afar</u> | <u>HOromo</u> | <u>Iraqw</u> | <u>Beja</u> |
| SOV order                                |  | yes           | yes         | yes           | yes          | yes         |
| Word order inside the NP:                |  |               |             |               |              |             |
| N modifier                               |  | yes           | no          | yes           | yes          | yes/no      |
| Mod. occupy a single syntactic position: |  |               |             |               |              |             |
| a. free order of mod.                    |  | <u>yes</u>    | no          | no            | no           | no          |
| b. multiple mod. --> coordination        |  | <u>yes</u>    | no          | no            | no           | no          |

This brief excursion into the noun modification in other Cushitic languages shows that Somali is unique in that the modifiers occupy a single position: all other languages have multiple slots for different types of modifiers, as can be seen from the strict ordering among them in Afar, Harar Oromo and Iraqw. In Beja definite adjectives and genitives even occupy different syntactic positions (pre- and post-nominal respectively).

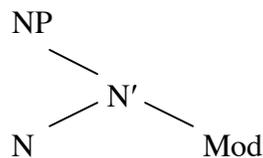
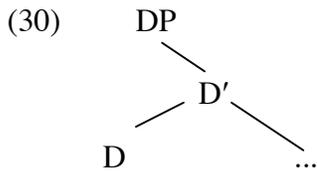
The cross-linguistic data also show that the properties of Somali noun modification cannot be reduced to variation in terms of headedness: Harar Oromo and Iraqw show the same fundamental word order properties (SOV, N>>modifier) while at the same time these languages distinguish several syntactic positions for the different types of noun modifiers.

**3. The structure of the Somali noun phrase**

To account for the properties of Somali noun modification we need to assume a structure that provides a single surface position for autonomous modifiers. This position hosts co-ordinations of NPs, relative clauses and adjectives, suggesting that it is a position for maximal projections.

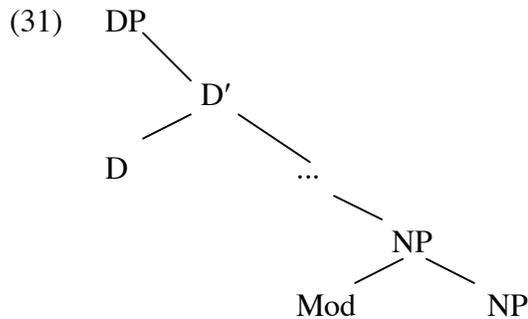
If we now assume that the modifiers occupy a DP-internal position, we may think of three possibilities.

The first possibility is that the modifiers occupy a complement position under N as in (30).



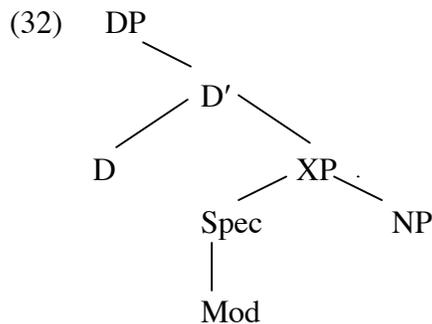
This structure is problematic, however, since it is traditionally assumed that there is a relationship of selection between a head and its complement, and there is no such selectional relationship between a N and its adjectival or relative clause modifiers.

A second possibility consists in a structure where the modifier complex is analysed as an adjunct to NP as in (31).



Since adjunction can in principle be iterated, this line of analysis has to postulate a separate ban on multiple adjunction within the Somali DP to account for the fact that the modifiers occupy a single structural position.

A third possibility consists in a structure where the modifier complex occupies a specifier position. This would derive the uniqueness of the modifier position.



Under the assumption that Spec is to the left of the head, modifiers cannot be in SpecDP because modifiers are on the right of the N+D complex they modify. This Spec has to be the Spec of an intermediate projection XP; so we would have to assume an additional functional projection (XP) for which we have no supplementary evidence.

We have no empirical argument to decide for or against these three options. Faced with the problems they raise, however, we would like to propose a different hypothesis.



**References**

- Almkvist, H. 1881. *Die Bischari-Sprache Bd 1*. Uppsala.
- Bliese, L. 1977. A generative grammar of Afar. Ph.D. diss., University of Texas at Austin.
- Greenberg, J.H., ed. 1966. *Universals of Language*. 2nd edition. Cambridge, MA: MIT Press.
- Hayward, D. 1998. Qafar. In A. Spencer and A. Zwicky, eds., *Handbook of Morphology*, 624-647. Oxford: Blackwell.
- Lecarme, J. 1996. Tense in the nominal system: The Somali DP. In J. Lecarme, J. Lowenstamm, and U. Shlonsky, eds., *Studies in Afroasiatic Grammar*, 159-178. The Hague: Holland Academic Graphics.
- Mous, M. 1993. *A Grammar of Iraqw*. Hamburg: H. Buske.
- Owens, J. 1985. *A Grammar of Harar Oromo*. Hamburg: H. Buske.
- Saeed, J.I. 1998. *Somali*. Amsterdam: J. Benjamins
- Serzisko, F. 1984. *Der Ausdruck der Possessivität im Somali*. Tübingen: G. Narr.

Sabrina Bendjaballah  
CNRS - UMR 8528 SILEX  
Université Lille 3  
BP 60149  
59653 Villeneuve d'Ascq Cedex  
FRANCE

sabrina.bendjaballah@univ-lille3.fr

Patricia Cabredo Hofherr  
CNRS - UMR 7023 / Paris 8  
c/o SAT Sciences du langage  
2 rue de la Liberté  
93526 – Saint-Denis cedex 02  
FRANCE

cabredo@ccr.jussieu.fr

## Syntactic Ergativity in Light Verb Complements

BALTHASAR BICKEL and JOHANNA NICHOLS<sup>1</sup>

*University of California, Berkeley*

### 1. Unexpected Syntactic Ergativity

Complementation is generally claimed to be the least likely construction to involve syntactically ergative pivots (Croft 1991, Dixon 1994, Manning 1996, Givón 1997, etc.). However, there is systematic syntactic ergativity precisely in complements in languages of the Kiranti family of Sino-Tibetan (Himalayas) and the Nakh-Daghestanian family (Caucasus). In these languages, certain verbs of modality, phase, and cognition agree with the S or O argument of their infinitive complement. Examples from Belhare (Kiranti, Sino-Tibetan; Nepal) are in (1):<sup>2</sup>

- (1) a.  $\emptyset$  khoŋ-ma nui-ka.  
[S] play-INF may-NPT:2[SG]S  
'You may play.'
- b. [ $\emptyset$   $\emptyset$  lu-ma] nui-ka.  
[A] [O] tell-INF may-NPT:2[SG]S  
'You may be told.' or 'I/S/he may tell you.'  
*Impossible*: 'You may tell him/her.'

The modal verb agrees in person with the S (intransitive subject) in (1a), the O (transitive object) in (1b). We will provisionally call this AGREEMENT CLIMBING. Agreement Climbing is SYNTACTICALLY, not morphologically, ergative. (2) shows Belhare verb agreement (also cf. Bickel 2001, in press). Agreement markers do not align S with O (or with A for that matter). The alignment patterns for all regular verbs, including *nus-* 'may', are variously three-way or neutral, depending on person, but nowhere ergative: first person agreement (underlined in (2) for clarity) is three-way; second person agreement is neutral (boldface in (2); the *-ga* ~ *-ka* alternation is morphophonemic); third person agreement is three-way, with a trace of accusative alignment (compare zero-marked S forms with zero-marked A forms when the O is third person).

<sup>1</sup> For contact information see the authors' website: <http://www.uni-leipzig.de/~autotyp>

<sup>2</sup> Apart from *nus-* 'may', another modal verb, *khes-* 'must' shows the same behavior, but also allows for impersonal constructions (similar to French *il faut*).

(2) Belhare verb agreement (singular number only; non-past forms)

|      | 1sgO                              | 2sgO                           | 3sgO                              | intransitive (S)          |
|------|-----------------------------------|--------------------------------|-----------------------------------|---------------------------|
| 1sgA |                                   | <i>luiʔ-na</i><br>tell-1sg>2sg | <i>luit-u-ŋ</i><br>tell-3sgO-1sgA | <i>nuiʔ-ŋa</i><br>may-1sg |
| 2sgA | <i>ka-lui-ka</i><br>1sgO-tell-2sg |                                | <i>luit-u-ga</i><br>tell-3sgO-2sg | <i>nui-ka</i><br>may-2sg  |
| 3sgA | <i>mai-luyu</i><br>1sgO-tell      | <i>n-lui-ka</i><br>3A-tell-2sg | <i>luit-u</i><br>[3sgA-]tell-3sgO | <i>nuyu</i><br>[3sgS-]may |

In the Chechen examples in (3), the gender prefix of the verb *v.ieza*, *j.ieza* ‘should’ is in agreement with the S/O of the infinitive: S in (3a), O in (3b). (V and J are genders.)

- (3) a. Muusaa  $\emptyset$  c'a-v.agma v.ieza  
Musa (V) [S] home-V.go.INF V.should  
‘Musa should go home.’
- b. Muusaas  $\emptyset$  disertaacii jaaz-j.a~ j.ieza  
Musa.ERG [A] diss.(J).NOM write-J.AUX.INF J.should  
‘Musa should write (his) dissertation.’

Thus, in Chechen as in Belhare, Agreement Climbing is ergative. Furthermore, the ergativity is truly SYNTACTIC, not morphological. Now, only nominative can trigger agreement in Chechen and Ingush. But that fact does not make Agreement Climbing MORPHOLOGICALLY ergative. There are two arguments showing that the ergativity is syntactic:

(i) Only those nominatives that are in S/O function can trigger Agreement Climbing. (4) shows an absolutive A in an accusative valence pattern (4a, b) with its causative (4c). The O of the causative infinitive (4c), but not the A of the non-causative (4b), can control Agreement Climbing. (Genitive case of *Muusaan* is assigned by *d.ieza* ‘must’; see below.) This shows that Agreement Climbing is not mechanically with a nominative, but with an S/O (provided the S/O is nominative).

- (4) Chechen *qiera* ‘fear’ < NOM LATIVE > (Ingush likewise)
- a. Muusaa sox qoeru  
Musa.NOM 1SG.LAT fear.INF  
‘Musa is afraid of me.’
- b. \*Muusaan [  $\emptyset$  sox qiera ] v.ieza  
Musa.GEN [A] 1SG.LAT fear.INF V.must  
‘Musa is supposed to fear me.’
- c. "Muusaan [  $\emptyset$  Ahwmad sox qieriita ] v.ieza  
Musa.GEN [A] Ahmed.NOM 1SG.LAT fear-CAUS.INF V.must  
‘Musa must instill fear of me in Ahmed. M must make A fear me.’  
[Semantically odd but syntactically OK.]

## *Syntactic Ergativity in Light Verb Complements*

(ii) Chechen and Ingush have the morphological means to assign nominative case to an A if the syntax ever sought an A controller. The auxiliary verb regularly assigns subject cases in Nakh, and progressive auxiliaries change an oblique subject case to nominative (see Conathan 2001). (5a) from Ingush shows that ‘read’ is transitive. (5b) shows that the progressive auxiliary ‘be’ assigns nominative (not ergative) case to the subject:

- (5) a. Muusaaz haara denna kinashjka diesh.  
Musa.ERG every day.DAT book.NOM D.read  
‘Musa reads a book every day.’  
b. Muusaa kinashjka dieshazh vy.  
Musa.NOM book.NOM D.read V.be.PRES  
‘Musa is reading a book.’

The progressive could surely function as an antipassive if the syntax required antipassivization. Evidently, therefore, the syntax of Chechen-Ingush infinitive complementation is uncoercedly ergative.

Apart from this syntactic ergativity of agreement climbing in complementation, Belhare has syntactic ergativity only in internal-head relativization (Bickel 2001 and in press), but in no other construction. Chechen and Ingush seem to have none at all.

### **2. Common Assumptions about the Distribution of Syntactic Ergativity**

These facts challenge common assumptions in the literature, indicated for example in the following statements:

Whenever Secondary concepts of the first variety [= predicates like ‘can’, ‘might’, ‘not’, ‘begin’, ‘finish’, ‘continue’, ‘try’ — BB&JN] are realized as lexical verbs, taking an object complement clause construction which involves another verb, the two verbs must have the same subject (S or A) irrespective of whether the language is accusative or ergative at morphological and/or syntactic levels. (Dixon 1994:135)

The control of Equi-NP-deletion (or coreference) in complement clauses is one of the least likely behavior-and-control features to show ergative-absolutive control. (Givón 1997:35)

A hierarchy of contexts for decreasing likelihood of syntactic ergativity, drawn from the literature, is shown in (6). Supporting subhierarchies are summarized in (7). (We are not aware of any empirical test of this hierarchy. We are beginning such a survey now as part of the AUTOTYP research program.)

- (6)
- 1: verb agreement
  - ▼
  - 2: relativization, focalization, interrogation, quantifier launch
  - ▼
  - 3: gapping in chaining and chained purposive constructions
  - ▼
  - 4: coreference marking in chaining, nonfinite complementation, reflexivization

- (7) Examples (square bracketing means ‘not applicable’):
- 1+2: Jacaltec (relativization/interrogation on S or O only; Van Valin 1981)
  - [1+] 2: Plains Cree (quantifier launcher must be S/O; Dahlstrom 1986)
  - [1+] 2+3: Dyrbal (S/O pivot in chaining and purposives; Dixon 1972)

Motivations, both formal and functional, for the hierarchy in (6) have been suggested in the literature. Possible functional motivations include the tight O-Verb relation (responsible for VP’s, idioms, etc.) motivating S/O control of agreement; relativization is most common on S and O (Fox 1987), and this motivates S/O control of focalization, interrogation, and other things commonly derived from or parasitic on relativization; cross-clausal coreference rests on topic continuity, and topics are mostly S/A (Givón 1983). A possible formal motivation (Manning 1996) is that control and reflexivization refer to a(argument) structure, i.e. <S> or <A, U> with A[ctor] always higher than U[ndergoer], while all other constructions refer to f(unctional) structure, e.g. <PIVOT, COMPL> with variable a-to-f mapping.

However, Belhare and Chechen-Ingush infinitive complementation reverse this hierarchy, showing syntactic ergativity in complementation and nowhere, or almost nowhere, else. This reversal is in conflict with all theoretical claims we are aware of, whether formal or functional.

### 3. Toward an Explanation

The key to understanding this unexpected pattern of syntactic ergativity comes from the observation that the matrix verbs involved are light verbs with a partly underspecified argument structure (‘arg-structure’): while their A-role must be filled by a referent (and is often thematically specified by the light verb itself), these verbs also open up a semantically unrestricted object (‘O’) slot which can be filled by a propositional expression (of the type *she wants that they do this*), but whenever this is possible, the slot is preferably filled by an element contained in the arg-structure of a dependent (subcategorized) verb.<sup>3</sup>

In the following, ‘lower verb’ or ‘lower arg-structure’ refers to the dependent verb, whose arg-structure provides the filler of the O-slot in the light verb. Lower

---

<sup>3</sup> The formal mechanism involved is best explicated as ARGUMENT COMPOSITION (cf., among others, Monachesi 1998 and Melnik 2000), but we leave elaboration of this point for another occasion. Here we focus on the typological facts that give rise to syntactic ergativity.

arg-structures are notated by lower case role labels (e.g. <a,o>), light verb arg-structures by upper case role labels (e.g., <A,O>). Cross-linguistically common instances of partly underspecified light verbs have phasal (*begin, stop, continue* etc.), modal (*can, must, may, etc.*), or cognitive (*want, see, think, etc.*) semantics.

There are two ways in which the lower arg-structure can satisfy the light verb's O-slot: by arg-structure unification, or by long-distance agreement. Both these options have particular lexical implementations that give rise to syntactic ergativity in complementation.

### 3.1. Argument Structure Unification

Unification of arg-structures results in clause union (complex predicate formation) and implies a single set of grammatical relations. There are many ways in which this can be achieved constructionally, e.g. through causative-style clause union that adds a grammatical relation, but the construction relevant for our purposes involves identification of the lower s or a-argument with the light verb's A-argument.

While the A-argument in the light verb is identified with the lower s or a-argument, the underspecified O-slot can be filled by a lower argument if there is one. This can occur only if the lower verb is transitive and thus contains an o-argument.<sup>4</sup> This is schematically explicated in (8), where the underspecified O-slot is represented by an underscore:

$$(8) \text{ Syntax:} \quad \begin{array}{ccc} & \text{A} & \text{O} \\ & | & | \\ \text{Arg-structure:} & \langle \text{a,o} \rangle + \langle \text{A, \_} \rangle & = \langle \text{A(=a), O(=o)} \rangle \end{array}$$

If the lower verb is intransitive, however, its only argument ('s') is already identified with the A-argument, and there is no other lower argument available that could fill the O-slot (cf. Haspelmath 1999 for a similar observation):

$$(9) \text{ Syntax:} \quad \begin{array}{ccc} & \text{A} & \text{O} \\ & | & | \\ \text{Arg-structure:} & \langle \text{s} \rangle + \langle \text{A, \_} \rangle & = \langle \text{A(=s), \_} \rangle \end{array}$$

As long as the the light verb has transitive syntax, the empty O-slot still needs to be satisfied. The only available candidate for this is the lower verb (or VP) as a whole, and this is indeed what characterizes a typologically common kind of Agreement Climbing construction.

This kind of Agreement Climbing is found, for example, in Spanish (and some other Romance languages). In Spanish, the O-slot of light verbs is generally satisfied through a proclitic in agreement with the lower o-argument. The lower a-argument is identified with the A-argument (10a):

---

<sup>4</sup> In ditransitives, the o-argument is the argument that maps into the direct or primary object.

- (10) a. (Lo=)quiero                      ver    a    Juan.  
           **3SG.M.ACC**=want.PRES.1SG    see.INF **ACC** J.  
           ‘I want to see Juan.’  
       b. (\*Lo=)quiero                      dormir.  
           **3SG.M.ACC**=want.PRES.1SG    sleep.INF  
           ‘I want to sleep.’

Agreement Climbing as in (10a) is generally optional, but in some dialects, e.g. in Argentinian Spanish, it is virtually obligatory (Fernando Zuñiga, p.c.). As shown by (10b), lower intransitive verbs cannot provide fillers of the O-slot. In this case, it is, as predicted, the VP (*dormir* ‘sleep’) as a whole that satisfies the O-slot. This does not result in an agreement-indexing proclitic (*lo=* ‘him’) because only specific nominal referents can trigger object agreement in Spanish.

A similar pattern is found in Basque (Iraide Ibarretxe, p.c.), but in this language, lower intransitive VPs are registered as third person singular objects by the agreement morphology of the light verb. This is shown by (11a); (11b) exemplifies O-agreement with the lower o-argument of a transitive dependent verb:

- (11) a. (Nik)    etxe-ra    etorr-i    nahi    dut.  
           1SG.ERG house-ALL come-PERF    want    **3SG.O**:PRES:1SG.A  
           ‘I want to come home’  
       b. (Nik)    liburu-a-k                      eros-i    nahi    ditut.  
           1SG.ERG book-DET-**PL[-NOM]**    buy-PERF    want    **3PL.O**:PRES:1SG.A  
           ‘I want to buy the books’

Essentially the same pattern of Agreement Climbing as in Spanish or Basque is found in a typologically widespread variety of languages, including Daghestanian (e.g., Haspelmath 1999 on Godoberi) and Indo-Aryan languages (e.g., Butt 1993 and Bickel & Yadava 2000 on Hindi). It is also found in Chechen and Ingush, where it characterizes modal verbs, e.g. Chechen:

- (12) a. Muusaan    disertaacii    jaaz-j.a~                      j.ieza  
           Musa(V).GEN    diss. (J).NOM write-J.AUX.INF    J.should  
           ‘Musa has to write a dissertation.’  
       b. Muusaan    c'a-v.gha    d.ieza  
           Musa(V).GEN    home-V.go.INF    D.should  
           ‘Musa must go home.’

The Chechen light verb *d.ieza* ‘should, must’ shows O-agreement with a lower o-argument, but if the lower verb is intransitive, the light verb shows default D-gender agreement.

*Superlight verbs and transitivity agreement*

In all of the instances surveyed in (10) – (12), the light verb has a fixed syntactic valence: it is always inflected transitively and governs the same transitive case

*Syntactic Ergativity in Light Verb Complements*

frame regardless of whether the lower verb is transitive or intransitive. However, some languages, have SUPERLIGHT verbs which assimilate in valence to the lower verb. Superlight verbs allow extension of Agreement Climbing to lower intransitive verbs:

- (13) a. Syntax:  $\begin{array}{c} \text{A} \quad \text{O} \\ | \quad | \\ \langle a, o \rangle + \langle A, \_ \rangle = \langle A(=a), O(=o) \rangle \end{array}$
- Arg-structure:  $\langle a, o \rangle + \langle A, \_ \rangle = \langle A(=a), O(=o) \rangle$
- b. Syntax:  $\begin{array}{c} \text{S} \\ | \\ \langle s \rangle + \langle S \rangle = \langle S(=s) \rangle \end{array}$
- Arg-structure:  $\langle s \rangle + \langle S \rangle = \langle S(=s) \rangle$

If the lower verb is transitive, as in (13a), the superlight verb has a transitive valence, and arg-structure unification proceeds just as with ordinary light verbs. If the lower verb is intransitive, as in (13b) however, the superlight verb has an intransitive arg-structure and an intransitive case syntax. The lower s-argument is now identified with S and can trigger S-agreement. The underspecified O-slot is removed.

Superlight verbs are found in Basque, Belhare, Chechen and Ingush. A Basque example is the verb *ahal* ‘can, be possible’; other modal verbs such as *behar* ‘must’ behave alike in the spoken language, but transitivity agreement is proscribed by the Academy (Iraide Ibarretxe, p.c.):

- (14) a. (Nik) liburu-a-k eros-i ahal ditut.  
 1SG.ERG book-DET-PL[-NOM] buy-PERF can 3PL.O:PRES:1SG.A  
 ‘I can buy the books.’
- b. (Ni) etxe-ra etorr-i ahal naiz.  
 1SG.NOM house-ALL come-PERF can 1SG.S:PRES  
 ‘I can come home.’

If the lower verb is transitive, as in (14a), the superlight verb is inflected transitively, and just as with ordinary light verbs, its O-slot is satisfied by the lower o-argument (‘books’) through agreement. But if the lower verb is intransitive, as in (14b), it is inflected intransitively and has intransitive case syntax (unlike with ordinary light verbs, cf. (11a)). As predicted by (13b), the S-function realizes the lower s-argument.

The same type of superlight verbs is found in Belhare. The verbs *lapma* ‘be about to’ and *hima* ‘already V, be able to V’ have this syntactic behavior:<sup>5</sup>

<sup>5</sup> See Bickel (2001, in press) for further discussion of infinitival complements in Belhare.

- (15) a. unna han lu-ma n-lapt-he-ga i?  
 3SG.ERG 2SG.NOM tell-INF 3[SG]A-be.about.to-PT-2[sg.O] Q  
 ‘Was s/he about to tell you?’
- b. unchik ta-ma n-lap-yu /\*n-lap-t-u.  
 3NSG.NOM come-INF 3NSG.S-be.about.to-NPT 3NSG.A-about.to-NPT-3[SG]O  
 ‘They are about to come.’

As shown by the ungrammaticality of the transitively inflected form *nlaptu* in (15b), valence agreement is obligatory. The resulting S-agreement marker realizes the lower s-argument, and the case on the subject is nominative, as predicted.

*Transitivity deponence and syntactic ergativity*

The syntax observed in (14) and (15) is nearly ergative: the superlight verb agrees with the lower o or s-argument. It is not fully ergative because the agreement markers are formally different: O-agreement in the case of a lower o-argument, and S-agreement in the case of a lower s-argument. However, full-fledged syntactic ergativity emerges if transitivity agreement is morphologically invisible, i.e. if it has only syntactic effects but no concomitant effects in verb morphology. This arises if the superlight verb has DEPONENT morphology. By deponent morphology (or deponence *tout court*) we understand any lexically marked mismatch of morphological and syntactic properties (see Corbett 2000, Bickel & Nichols, in press). The classic example of deponence is Latin verbs with transitive active syntax that lack active inflectional morphology and have only passive inflectional paradigms. Similarly, deponent superlight verbs are verbs with transitive or intransitive syntax that lack the corresponding inflectional morphology.

Belhare has both all-transitive deponent superlight verbs, which lack intransitive paradigms but show transitivity agreement in the syntax, and all-intransitive deponent superlight verbs, which lack transitive paradigms but show transitivity agreement in the syntax. The following illustrates one of five superlight verbs that are all-transitive deponents (the others are *talokma* ‘begin’, *manma* ‘finish’, *munditma* ‘forget’, and *nima* ‘know’):

- (16) a. unna han lu-ma n-tog-he-ga.  
 3SG.ERG 2SG.NOM tell-INF 3[SG]A-can-PT-2[SG.O]  
 ‘He had a chance to tell you.’
- b. unchik ta-ma n-tou-t-u.  
 3NSG.NOM come-INF 3NSG.A-can-NPT-3[SG]O  
 ‘They can come.’

(16a) illustrates the transitive version of the superlight verb, and the syntax is exactly the same as with regular superlight verbs (cf. (15a)): the lower o-argument triggers O-agreement in the superlight verb. In (16b), the superlight verb *tokma* ‘can, have the opportunity to’ agrees in transitivity with the lower intransitive verb, and this has the syntactic effects that the lower s-argument can trigger



which this is possible and in which the lower *o* can fill the O-slot in an all-intransitive deponent verb is through the mapping shown in (19).

The result of this, however, is syntactic ergativity: the lower *s*-argument in (18a) triggers exactly the same kind of agreement in the superlight verb as the lower *o*-argument in (18b).

Ancillary evidence for this analysis is twofold. First, the case marking found in (18b) confirms that the syntax is transitive despite the intransitive morphology of the superlight verb: there is no other situation in Belhare where intransitively inflected verbs combine with an ERGATIVE-NOMINATIVE case frame. Likewise, in (16b), the case frame is intransitive despite the transitive morphology, and again, there is no other situation in Belhare where a NOMINATIVE-only case frame could co-occur with a transitively inflected verb. It is only in these lexically marked instances of transitivity deponence that there can be a mismatch between case and agreement morphology.

Second, deponence is an independently attested phenomenon in Belhare and many of its sister languages. Experiential predicates like *khikma* ‘taste bitter’, for example, are all-intransitive deponents that have two syntactic actants, an experiencer in A-function and a stimulus in O-function (see Bickel 1999 and in press for discussion of the syntactic properties of these actants). Belhare’s sister language Limbu has experiential predicates that are all-transitive but that have an intransitive syntax (Michailovsky 1997).

The analysis proposed for Belhare essentially carries over to Ingush and Chechen. There are regular superlight verbs undergoing transitivity agreement. Transitive infinitives may and preferably do trigger transitive auxiliaries in the superlight verb; intransitive infinitives must have intransitive auxiliaries (cf. Melnik 2000 for Chechen):

- (20) a. Laqa juola-jyr.  
           play.INF J.start-J.TRANS.AUX.WP  
           ‘She started playing [the instrument (J-gender)].’ (7D)
- b. Chaarx c'eaxxaa qesta juola-jalar.  
           wheel(J) suddenly turn.INF J.begin-J.INTR.AUX.WP  
           ‘The wheel suddenly started turning.’

With a transitive lower verb, as in (20a) the superlight verb is inflected intransitively (on its auxiliary) and agrees with the *O=o* actant, following the pattern in (13a). If the lower verb is intransitive, as in (20b), the superlight verb is intransitive as well and agrees with the lower *s*-argument (‘wheel’ (J gender)).

Chechen and Ingush also have a few superlight verbs that are similar to Belhare superlight deponents. One of them, *d.ieza* ‘should’, is illustrated in the Chechen example (3) above. Unlike the superlight verbs (about 4 in total) illustrated for Ingush in (20), these deponent verbs have the same effect of syntactic ergativity as we observed in Belhare: syntactically, they assimilate in transitivity,

whence agreement is with the lower s- or o-argument; but morphologically, this is invisible, and the agreement morphology always takes the same form. However, unlike in Belhare, the transitive vs. intransitive distinction is not an obligatory property of Ingush morphology. It is only apparent if verbs take an auxiliary, as they do in (20). The verb in (3a) is a simplex verb. Therefore, the reason why this verb does not manifest morphologically visible transitivity agreement is that it does not take an auxiliary and not that it lacks parts of a paradigm. We call this a QUASI-DEPONENT.

And, as in Belhare, transitivity deponence is found independently in other parts of the Ingush (and Chechen) lexicon as well. The Ingush verbs *qeika-d.u* ‘cough’ and *qoa-d.u* ‘find time, manage’ are always inflected transitively (i.e. take transitive auxiliaries), yet they have intransitive valence with one single actant.

### 3.2. Long-Distance Agreement

In the scenario described so far, the restriction of matrix O agreement to lower o arguments results from the fact that the lower s- or a-argument is coreferential with the higher A argument, hence unavailable as filler of the O-slot in the light or superlight verb. However, some Kiranti languages (e.g., Belhare) and some Nakh-Daghestanian languages (e.g., Tsez) extend the same type of light verb complementation, with the same type of matrix agreement, to disjoint s/a-arguments. The result is yet another complementation construction with a syntactically strictly ergative pivot.

In these cases, clauses are not unified but preserve two distinct sets of grammatical relations. However, the empty object slot in the light verb is satisfied by an argument from the lower clause through LONG-DISTANCE AGREEMENT (Polinsky & Comrie 1999). Example (21) illustrates this for Tsez:

- (21) a. eni-r            [už-ā    magalu            b-āc'-ru-λi]            b-iy-xo.  
           mother-DAT [boy-ERG bread.III.NOM III-eat-PT.PART-NMLZ] III-know-PRES  
           ‘The mother knows the boy ate the bread.’
- b. eni-r            [uži            ø-āy-ru-λi]            ø-iy-xo.  
           mother-DAT [boy.I.NOM I-arrive-PT.PART-NMLZ] I-know-PRES  
           ‘The mother knows that the boy arrived.’

As Polinsky & Potsdam (in press) show, the structure of these sentences is biclausal; yet it is the lower a (21a) or the lower s (21b) argument that is recruited to fill the O-slot in the light verb’s arg-structure through long-distance agreement. Note that in contrast to the clause union constructions discussed earlier, the lower a-argument *užā* ‘boy’ in (21a) is not integrated into the light verb’s arg-structure. It is only s and o arguments that can trigger long-distance agreement, and this is another source for syntactic ergativity in light verb complementation.

A similar construction is found in Belhare with the light verbs *konma* ‘want’ and *nama* ‘stop’:

- (22) a. khali [set-ma] ka-ŋ-koĩ-yu.  
 only kill-INF 1INCL.O-3NSG.A-want-NPT  
 ‘They just want us to get drunk.’ (lit., ‘they want [the beer] kill us.’)  
 b. [ten-ma] ma-ŋ-narend-he.  
 beat-INF 1SG.O-3NSG.A-stop-PT  
 ‘They stopped x from beating me.’ *or* ‘They stopped beating me.’

In these examples, the lower verb is transitive and its o-argument fills the light verb’s O-slot through agreement. The lower a-argument (that which ‘kills’ in (22a), the person(s) who beat in (22b)) is not syntactically identified with the light verb’s A-argument. This is again in contrast with the constructions involving arg-structure unification.

However, note that while the lower a-arguments in (22) are not SYNTACTICALLY identified with the higher A-arguments, they are left unexpressed, and in a suitable context, the lower a-arguments can be PRAGMATICALLY identified with the A-arguments (cf. the alternate translation of (22b)). Indeed, the most common usage of *konma* ‘want’ and *nama* ‘stop’ relies precisely on this interpretation. A likely reason for this is that Belhare infinitives do not allow antipassive formation, so that a syntactic constraint AGAINST a=A identification would make it impossible to have topic continuity in such ordinary clauses as ‘he wants to buy beer’ or ‘she stopped drinking beer’.

Thus, though a and A can be distinct, they most often are coreferential, and this assimilates the construction to arg-structure unification as discussed earlier. Belhare long-distance agreement constructions indeed seem to be developing into clause union constructions. Independent evidence for this comes from the fact that *konma* ‘want’ and *nama* ‘stop’ allow (partly) optional transitivity agreement, the hallmark of superlight verbs and arg-structure unification. Thus, along with the transitive version of *konma* ‘want’, intransitive inflection is possible too (23a). With lower intransitives, intransitive inflection is obligatory (23b):

- (23) a. lu-ma koĩ-yu.  
 tell-INF [3SG.S-]want-NPT  
 ‘He wants to tell.’ *or* ‘He<sub>i</sub> wants x to tell him<sub>i,j</sub>.’  
 b. ta-ma koĩ-yu / \*mai-koĩ-yu / \*koĩ-t-u.  
 come-INF [3SG.S-]want-NPT 1SG.O-[3SG.A-]want-PT [3SG.A]want-NPT-3SG.O]  
 ‘He wants to come.’ *or* ‘He wanted x to come.’

However, unlike with superlight verb constructions, the lower a or s-argument still does not have to be coreferential with the A-argument of *konma* ‘want’.

The asymmetric distribution of transitive and intransitive versions in (23) is the same as the one observed in Chechen and Ingush (see discussion of example (20) above), and results from the fact that all these languages are base-intransitive languages, where the intransitive forms are the default, and transitives the derived forms (Nichols *et al.* 1999).

#### **4. Conclusions and Implications**

The findings presented here suggest that, despite common assumptions to the contrary, true ergative syntax is possible in complementation structures. It is possible if a language has either (i) deponent superlight verbs that allow extension of argument structure unification to intransitives, or (ii) long-distance agreement. While in option (ii) there is no intrinsic need for syntactic ergativity to arise, option (i) appears to necessitate ergativity because, as we showed in Section 3.1, the unification construction that deponent superlight verbs rely on have an intrinsic bias to take the lower o or s arguments as fillers of the superlight verb's empty O-slot. The a-argument is unavailable because it is identified with the superlight verb's A-argument.

Thus, type (i) syntactic ergativity is the result of specific lexical properties in argument structure (light vs. superlight verbs) and morphology (deponence). This suggests that universal hierarchies of the likelihood of syntactic ergativity may be successfully relativized to lexical properties.

#### **Abbreviations**

2SG second person singular; 3PL third person plural, etc; I, III Tsez genders; A subject of transitives; ACCusative; ALLative; AUXiliary; CAUSative; DATive; DETerminer; ERGative; GENitive; INFinitive; J a Chechen-Ingush gender; LATive; M masculine gender;; NMLZ nominalizer; NOMinative; NPT nonpast; NSG nonsingular; O (primary or direct) object of transitives; PARTiciple; PERFect;; PRESent; PT past; S single argument of intransitives; TRANSitive; V a Chechen-Ingush gender; WP witnessed past. Elements in square brackets are features expressed by zero morphemes; '=' denotes a clitic boundary.

#### **Acknowledgments**

NSF 96-16448 (JN, PI) (Chechen, Ingush). Arbi Vagapov (Chechen examples). SNSF 8210-053455 (BB, PI) (Belhare). Lekh Bāhadur Rāī (Belhare examples).

#### **References**

- Bickel, B. 1999. Grammatical relations, agreement, and genetic stability. Ms., University of California, Berkeley.
- Bickel, B. 2001. Belhare. In G. Thurgood and R. J. LaPolla, eds., *The Sino-Tibetan languages*. London: Curzon Press.
- Bickel, B. In press. Hidden syntax in Belhare. In G. van Driem, ed., *Himalayan Linguistics*. Berlin: Mouton de Gruyter.
- Bickel, B., and J. Nichols. In press. Inflectional morphology. In T. Shopen, ed., *Language Typology and Syntactic Description*. Cambridge: Cambridge University Press (revised second edition).
- Bickel, B., and Y. P. Yadava. 2000. A fresh look at grammatical relations in Indo-Aryan. *Lingua* 110: 343-73.

- Butt, M. 1993. A reanalysis of long distance agreement in Urdu. In J. Guenter, B. Kaiser, and C. Zoll, eds., *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistics Society*, 52-63. Berkeley, CA: Berkeley Linguistics Society.
- Conathan, L. 2001. Ergative and binominative constructions in Ingush. Ms., University of California, Berkeley.
- Corbett, G. G. 2000. Four inflectional phenomena and the notion 'possible lexical item'. Paper presented at the Northwest Slavic Linguistics Conference, Berkeley, CA.
- Croft, W. 1991. *Syntactic Categories and Grammatical Relations*. Chicago: University of Chicago Press.
- Dahlstrom, A. L. 1986. Plains Cree morphosyntax. Ph.D. diss., University of California, Berkeley.
- Dixon, R. M. W. 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Fox, B. 1987. The noun phrase accessibility hierarchy reinterpreted: Subject primacy or the absolutive hypothesis? *Language* 63: 856-70.
- Givón, T., ed. 1983. *Topic Continuity in Discourse*. Amsterdam: John Benjamins.
- Givón, T. 1997. Introduction. In T. Givón, ed., *Grammatical Relations*, 1-84. Amsterdam: John Benjamins.
- Haspelmath, M. 1999. Long-distance agreement in Godoberi (Daghestanian) complement clauses. *Folia Linguistica* 33: 131-51.
- Manning, C. D. 1996. *Ergativity: Argument Structure and Grammatical Relations*, Stanford, CA: CSLI.
- Melnik, N. 2000. Agreement climbing in Chechen: An HPSG analysis. Ms., University of California, Berkeley.
- Michailovsky, B. 1997. Catégories verbales et intransitivité duale en Limbu. *Studi Italiani di Linguistica Teoretica e Applicata* 26: 307-25.
- Monachesi, P. 1998. Italian restructuring verbs: A lexical analysis. In E. Hinrichs, A. Kathol, and T. Nakazawa, eds., *Syntax and Semantics Vol. 30: Complex Predicates in Nonderivational Syntax*, 313-68. New York: Academic Press.
- Nichols, J., D. A. Peterson, and J. Barnes. 1999. Causativizing and decausativizing languages. Association for Linguistic Typology Meeting, Amsterdam.
- Polinsky, M., and B. Comrie. 1999. Agreement in Tsez. *Folia linguistica* 33: 4-25.
- Polinsky, M., and E. Potsdam. In press. Long-distance agreement. *Natural Language and Linguistic Theory*.
- Van Valin, R. D., Jr. 1981. Grammatical relations in ergative languages. *Studies in Language* 5: 361-94.

## The Intersubjectivity of ‘Self’ and ‘Other’ Tokens in Communication: Moving Between Represented Subjectivity and Conceptual Objectivity

MELINDA Y. CHEN

*University of California, Berkeley*

It is a fact both original and fundamental that these “pronominal” forms do not refer to “reality” or to “objective” in space or time but to the utterance, unique each time, that contains them.... The importance of their function will be measured by the nature of the problem they serve to solve, which is none other than that of intersubjective communication. Language has solved this problem by creating an ensemble of “empty” signs that are nonreferential, with respect to “reality”. (Benveniste 1966:219)

How is a person subjectified through language, as an ongoing repository for the endowment of subjectlike qualities? When, and for whom, does one’s own represented subjectivity slide into the “subjective objectivity” of an intersubjective reality? And what does it mean for this to happen linguistically?

This paper examines the interpretation of tokens of ‘self’ and ‘other’ with a focus on the interplay between cognitive semantics and (intersubjective) interaction, warns against imputing fixity to pronominal referents, and explores dynamics of meaning production vis-à-vis subjectivity. In the coming discussion, *I* and *you* are traced in a tradeoff of insults. Crucial is the idea that two minds cross over shared objects – as well as the fact that in an insult exchange, an assumption shared by the speakers is that viewpoints are heterogeneous rather than identical, and representations of selves may be at stake.<sup>1</sup> A mental spaces framework provides a way to represent not only conceptual representations that are the preeminent constituents of linguistic meaning, but diverse other aspects of interaction – intersubjective awareness, indexing, and identity types – without sacrificing systematicity.<sup>2</sup> My theoretical departure is Benveniste’s insightful work on

---

<sup>1</sup> Identity, so complex it often seems ungraspable, once grounded in cognition and seen as a matter of representation, serves as an excellent focal point for explorations of intersubjective meaning produced on-the-fly.

<sup>2</sup> Mental Spaces Theory (Fauconnier 1994, 1997, and others), or MST, is a general theory of conceptual meaning which postulates the dynamic change of conceptual structure in response to a conceptualizer’s interactions with entities “in the world”, whether or not this involves language

pronouns, which can be expanded and revised for a modern cognitive linguistics context.

Language is constitutionally intersubjective. Lexical meaning is never quite owned by one specific person, and yet such meaning is only human (rather than objective) in some crucial sense; hence that meaning is continuously shared among (linguistic) agents. It is also useful to note that these sharings are often the result of a power struggle; they must, sometimes literally, be negotiated, between or “inter-subjects”. While the language being negotiated is implicitly shared, the viewpoints around which it is negotiated may well be antagonistic. Let us then consider intersubjectivity in its complex nature: doing language involves shared communication between two neurally, culturally, and cognitively heterogeneous agents who will (whether consciously or not) be at odds in a number of ways.

If conservative about the givenness of information about an interlocutor, we might depict one such interlocutor’s mental space configuration for their conversational situation as in Figure 1 below.<sup>3</sup> This is a continually managed, intersubjective configuration of mental spaces, any of which might serve as a *focus space* and hence be modified in content. Such a multiplicity of spaces suggests there is room for all kinds of inferencing. At the same time, potentially adversarial conditions do not completely divest participants of collaborative meaning-making. This collaboration is partly manifested in the sense that there are inherent constraints to interpretation – such as the prioritized use of a particular set of mental spaces. That is, there may be interpretative complexity, but intersubjective awareness ensures a certain amount of cooperation. As for how to sort through this complexity, I explore later how perspectivization (indication of a particular viewpoint as a source for information) may be explicitly or implicitly signalled.

---

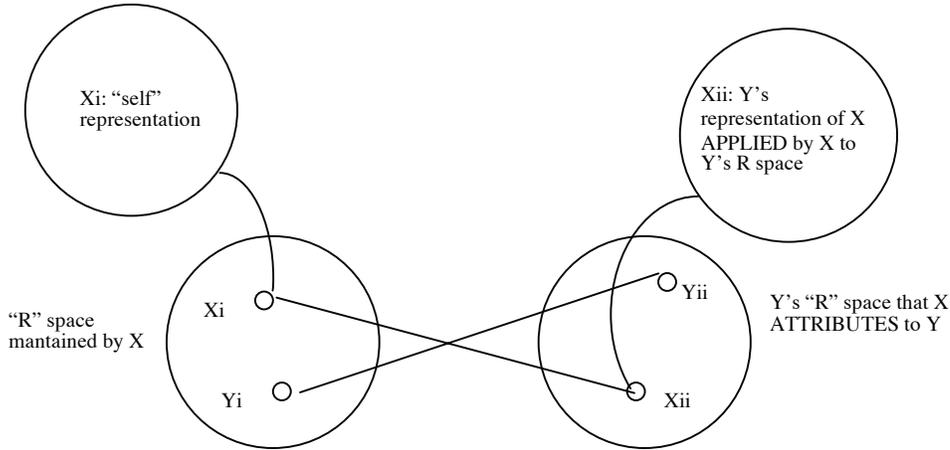
activity. MST is apt for integrating much information understood to be classically pragmatic, and has offered advantageous explanations for a number of classic pragmatic-semantic problems.

The mental space is a construct representing any packaged organization of information in cognition. Only distantly related to classical semantics’ “possible world”, it comprises a diversity of conceptual information, such as entities and frames. Informed by context, schematic or skeletal spaces become enriched and particularized. Conceptualizers at minimum maintain a “real” space R, in which the current interaction’s participants and relationships (if they are available) are installed. Upon this base, utterances constitute streams of cues for constrained conceptual development. Anaphoric or other pronominal references (*she, that cat*) either make use of existing correspondences between spaces’ elements, or generate new ones. Space builders are another type of cue: for instance, “is” revises a current, or Focus, space; “believes” creates a new Belief space; “was” and “before” create Past-Time spaces.

<sup>3</sup> My thanks go to Kevin Wiliarty for this lucid configuration, which is useful for the exposition. I am fully responsible for any errors.

*Intersubjectivity of ‘Self’ and ‘Other’ Tokens*

Figure 1: Intersubjectivity in Conversation: a Four-Space Representation of Interaction and Identity Maintained by One Interlocutor (X)



The following two hypothetical exchanges between a waiter and customer in an eating establishment differ just enough to illustrate a distinction. The conversational context is accounted for by a *RESTAURANT* frame, where a certain power configuration – which is relevant to the playing-out of collaborative, intersubjective meaning – is partly manifest in the roles of server and served, and in scripts for the proper interactive behavior of each participant. Because I lack the space to draw progressive mental spaces imagery, instead I record the conceptualization events prompted by explicit linguistic cues, as well as the inference production likely to result.

Exchange 1: Insult (without resistance).

| DIALOGUE                                                       | MENTAL SPACE EVENTS                                                                                                                                                                                               |
|----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| a = customer<br>b = waiter                                     | Ia a’s self-representation in space Ra;<br>Ib b’s self-representation in space Rb;<br>YOUa a’s representation of b in space Ra;<br>YOUb b’s representation of a in space Rb.<br>⇒ indicates inference production. |
| <b>(1) a: HEY, I said I didn’t want tomatoes on my burger!</b> | [Ia=said[Ia=didn’t want tomatoes on burger]]<br>⇒ Iapast=don’t want tomatoes                                                                                                                                      |
| <b>(2) b: I thought you wanted tomatoes.</b>                   | [Ib=thought[YOUb=wanted tomatoes]]<br>YOUbpastbelief=want tomatoes.<br>Iapast != YOUbpastbelief; two MS must be maintained.<br>Iapast: don’t want tomatoes<br>Ibpastbelief: YOUb=want tomatoes                    |

|                                                                                 |                                                                                                                                                                                                                                                                                                                                                                     |
|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>(3) a: Well, I DIDn't... [pause] you're obviously a numbskull.</b></p>    | <p>[Ia=didn't want tomatoes]<br/>reconfirms viewpoint MS-Ia, viewpoint a in a's R space (reconfirms R space)<br/>[YOUa=numbskull]<br/>from this viewpoint, depicts YOUa. YOUa has no conflict with Ib. If Ib's thought=mistake (customer is always right, at least more powerful), Ib "deserves" the label! Ib's belief != a's R space; Ib=wrong; Ib=numbskull.</p> |
| <p><b>(4) b: Hey, I'm sorry.</b></p>                                            | <p>[Ib=sorry]<br/>Ib=sorry.<br/>⇒ Ib concedes culpability.</p>                                                                                                                                                                                                                                                                                                      |
| <p><b>(5) a: Yeah... well, you can try doing your job next time, maybe.</b></p> | <p>[YOUa=can try doing job next time]<br/>⇒ YOUa doesn't do job</p>                                                                                                                                                                                                                                                                                                 |

Here A promulgates his own offensive representation of B. If power relationships are such that A is afforded more “meaning authority” in the local context (he is here), then both A and B know that A’s reality has extra purchase in some sense.<sup>4</sup> B’s own *I* representation may well differ from A’s (unless B already holds, or is susceptible to, negative self-representations), in which case it points to an alternative representation of *I*. In a more implicit case, A’s characterization of B paints A’s representation of B, and B’s defense paints B’s representation of B. These characterizations can be implicitly made.

In the second example – Exchange 1’s complement of a sort – instead of (4) above, we see active resistance on the part of the waiter in (9):

Exchange 2: Insult (with resistance)

|                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>(6)-(8): see (1)-(3) above</b></p>                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <p><b>(9) b: Dude, NUMBskulls are those people like yourSELF who avoid healthy foods.</b></p> | <p>(Redefines numbskulls to fit a)<br/>[numbskulls=avoid healthy foods]<br/>⇒ YOUb=eats hamburgers<br/>⇒ YOUb=avoids healthy foods<br/>⇒ YOUb=numbskull<br/>⇒ Ib eats healthy foods<br/>⇒ Ib is not a numbskull<br/>[Ib=NOT numbskull] now arouses conflict with YOUa. Mental spaces disassociate/de-link.<br/>[YOUb=numbskull; YOUb=unhealthy; Ib=eat healthy foods; Ib=NOT numbskull]<br/>MS for Ib is: Ib's thought=mistake; Ib=NOT numbskull; Ib=eat healthy foods;<br/>MS for YOUb=numbskull; YOUb=unhealthy</p> |

<sup>4</sup> On my use of “meaning authority”: The concept is certainly more complex than **to be monolithically granted to one interlocutor for the full course of the interchange. For more reading on such issues, see (Butler, *Excitable Speech*, and Bourdieu, *Language and Symbolic Power*).**

## *Intersubjectivity of ‘Self’ and ‘Other’ Tokens*

There are two ways in which the waiter can be understood to resist the application of the label *numbskull* (as well as, more generally, the exercise of power that accompanies the utterance). First, the waiter redefines *numbskull* by referring to a type that apparently includes his interlocutor (people who avoid healthy foods). This act also invites the inference that his own equation with the label *numbskull* must be incorrect. Until the customer has a chance to respond, the waiter has momentarily exercised a limited meaning authority.

Given that reality is not objective, it can be contested, and power helps determine whose representations win. In a conceptualist theory, reality is fundamentally treated as a matter of representation.<sup>5</sup> That is, for each language interpreter, a reality space *R* – again, recall, informed by schemas, idealized or stereotypical models, and details of the immediate context – functions as a base within which, and from which, meaning itself is constructed. But what informs *R* may well shift: Whose contribution wins is precisely what gets differing distributions in the two exchanges above. For instance, the resistive waiter wrests, in (9), a kind of authority over the shared portion of Reality space, as a consequence of the collaborative conversation (a la Grice) presumed to minimally exist, and set his terms – literally – upon it. Note that it is highly possible that a given *R* space is at times assumed to be shared – that is, that all participants possess in some sense the same “subjective objectivity”.<sup>6</sup> Objective facts or truths reach this status precisely because the group that shares these facts or truths concur on their meaning. I later discuss the ways that power relations, and variations in egocentric orientation, can affect what is inferred as objective.

How do interaction and semantics work together in the exchanges above? To review the technology of the pronouns themselves, Benveniste, in “The Nature of Pronouns”, distinguished the 1<sup>st</sup> and 2<sup>nd</sup> person pronouns (presumably true for all languages) from 3<sup>rd</sup> person pronouns in the following way: *I* and *you* mean no more and no less than that one who is either speaking or being addressed at the moment of the utterance (1966). Further, it is precisely *I* that allows an individual speaking subject into being. Thus, while the *I* and *you* pronouns seem to afford crucial means for the represented experience (and hence the enabling) of personhood and subjectivity, they also, almost paradoxically, index that same subject.

With Benveniste’s strict and entirely appropriate delimitations about the meanings of 1<sup>st</sup> and 2<sup>nd</sup> person pronouns in mind, I make below some specific claims regarding the possibilities of over-imputing fixity to the entities (language objects and referents alike) associated with the semantics of these pronouns. I address two possible erroneous inferences: first, that the interlocutors serving as

---

<sup>5</sup> I do not mean to imply that this representation exists devoid of embodied, dynamic engagement in the world.

<sup>6</sup> By subjective objectivity, I mean that both interlocutors are collaboratively compelled to impute a common Reality space, especially in a powered situation when the inferred perspective is automatically accepted as authoritative. Within this space a given element effectively reaches the status of objectivity, though its evaluation may well have subjective origins and can certainly be attributed to human cognators.

referents to the pronouns are essentially fixed; and second, that *I/you* point to unitary bodies or subjects. In reality, the interlocutors are both always present and being shaped at the time of speaking, partially by the speech; furthermore, the process of shaping is not exhaustively applied to “the subject”, but rather operates on selected aspects of subjectivity.<sup>7</sup>

It is easy enough to imagine an objectified, intellected body of information, separate from the interlocutors, which is being discussed. This is effectively to fix the interlocutors as constant, external elements, beyond the inclusive reach of regular conceptualization. What makes the fixity a tempting default for practitioners and conversationalists alike is the possibility of a kind of teleological overflow, outpacing the borders of what is definitional fixedness. It is only as indexical elements without objective content that *I/you* can fairly be seen to possess fixed semantics. Note, however, that even if they are purely indexical, *what* they index is potentially mobile, precisely because in the process of indexing the entity itself is being (conceptually) shaped. *I/you* do not simply refer the expression to an aspect of the context. They actively *create* that context by pointing in some sense to what needs to be modified. Thus a real indeterminacy lurks beneath the seeming fixity of *I/you* as index-functional lexical items. In sum, then, the limited semantics of the person deictics in no way ontologically secures what it says nothing about.

Also supporting the possible imputation of ontological fixity are interpreters' belief in object constancy, or, on a more particular human scale, identity permanence (defined as “the understanding that the same person is very different at different times in his or her life, but is nonetheless ‘the same’”, Sweetser and Fauconnier 1996), and the belief in a consistent self.<sup>8</sup> But given the insistent dynamism of what we understand consciously and unconsciously to be reality, constancy in meaning must exist alongside changes in other meaning. Applied to the interlocutors behind the pronouns, this means that the fluidity of the interlocutors' very complex identities themselves must also be taken into account. Recall that Benveniste has written that pronouns allow us to assign characteristics – subjectivity – to those involved in conversation; in fact, such is the precise means of achieving subjecthood (Benveniste 1966).

The *belief* in the durativity of the *I*, ironically, plays a role quite opposite to that of fixing the *I*. The *belief* in the durativity of the *I* is what enables the actual

---

<sup>7</sup> A double edge inheres in the concept of “ground”. Grounding might easily be thought as that fixed system of relatively intersubjectively stable reference points by which we may, somehow, fix or specify a conceptualization. But the degree which we may think of *I* and *you* as “given” aspects of the ground seems highly variable. The mental spaces and cognitive grammar frameworks both seem to have made presumptions of givenness from time to time, but *I* and *you* may only really be maximally fixed as part of an idealized, schematic conversational speaker/hearer template.

<sup>8</sup> The relation between object constancy and conceptual representation of the object, I suggest, may be analogized to the relation between an idealized cognitive model and the messier reality of actual linguistic behavior.

## *Intersubjectivity of ‘Self’ and ‘Other’ Tokens*

*meaning of I* to assume a gradual development, rather than enjoying life and death at every moment. It is what gives us an ongoing (conceptual) trace which is subject to modification. For the conceptualist view, then, Benveniste’s time-sliced reality of discourse (286), with validity (of the *use of I*) restricted to the moment, needs complementation with the belief that there is a consistency of self (of the *meaning of I*).

With regard to the possible inference about unitary referents, Ronald Langacker has taken the position that interlocutors are perhaps *always* an implicit presence in the conceptualizations, no matter the object. In discussing the interlocutors’ role in being the afforders of meaning, he argues that the figure of the speaking subject, S, inheres in some way (“at least peripherally”) in every element’s encyclopedic semantic value, whether or not this has been conventionalized in terms of a templatic subjectivity. And “at the very least there is always an implicit conceptualizing presence” (Langacker 1997). Thus an interlocutor is present in two ways: first, as explored above, as conceptualized individuals endowed with character by virtue of the use of the “pointing” pronouns; and second, as overseeing conceptualizers for every instantiation of meaning. While the degree of onstage-offstageness is in constant flux, there is no reason not to think that our ongoing cognitive tasks include “writing the right information to the right blend” – meaning that we must find ways to integrate (in the form of a cognitive blend) the overseeing-conceptualizer of the last utterance with the *you* of the present one, and so forth.

Each interlocutor, then, is itself a blend, between a subject whose conceptualizing capacity is profiled, and an individual continually embellished with the characteristics of the propositions in which the pronouns *I* and *you* have been embedded.<sup>9</sup> Theoretically, we have assumed that interlocutors do not implicitly appear in a conceptualization (see Goffman 1959, Sanders and Sporeen 1997). And yet *conceptual presence* is essential: language is what we do. This means that we are forever being conceptually shaped, however implicitly and in the background, beyond conscious awareness this may be.<sup>10</sup>

Now we must address the mechanisms of the conceptual shifting behind the imagined fixed self. There is a subtlety to what it is that *I* and *you* do, that moves beyond the simple creating-indexing duality. So far I have said that these pronouns serve as pointers to conceptualizations of interlocutors that are being

---

<sup>9</sup> It is also worth noting that the cognitive representation of the *interlocutor* and that of the participating *subject* necessarily coexist in a blend.

<sup>10</sup> The conceptualist account above suggests that deictics (including the *I/you* pronouns), though they do have a necessity requirement for reference to the immediate instance of discourse, cannot by nature be so *radically* distinct from other utterable constructions. It might be more reasonable to think of them as so radically installed at one end of a graded scale along a certain semantic dimension (in this case that they refer almost templatically to the local conversational context), that, for instance, “you” simply cannot be cited except in an explicitly citational construction (“She said ‘You scum’”), or, perhaps, only with dramatic help of prosody, emphasis, and other constructional disambiguators (perspectivizers).

enriched by the current discourse. Subjectivity itself, (as identity) is in some very real sense an indivisible and highly variable and individualized mass of subject qualities.<sup>11</sup> Would *I/you* then simply refer unvaryingly to this rather monstrous, gestalt unity: the pan-subjective representation of an interlocutor? From the perspective of the mind-in-action, while a seeming infinitude of interpretations of *I* are possible, it would be implausible for a conceptual instantiation to include all aspects of subjectivity at once, at equal degrees of salience. Rather, the conditions of salience and schematic representation must still apply. What are our alternatives, then?

One alternative is offered by the specificity of cognitive domains. In encyclopedic cognitive semantics, every lexical item is understood to evoke a number of possible relevant cognitive domains. There is no reason to exclude the “special” case of pronouns from this phenomenon. In this view, then, in context *I/you* should simply alert some array of relevant subjective qualities (judgment, attitude, emotion, physical being, spirit, and so on). But what reason have we to believe, besides applying a vague principle of cognitive economy, that subjectivity is organized in our representations such that we deal with subtypes, rather than wholes, of subjectivity?

The grammaticalization literature has demonstrated several structured ways that specific domains of subjectivity – and not others – are called up upon certain lexical appeals to the subject. For example, the verb *promise* can indicate either deontic (*Jerry promises*) or epistemic modality (*the economy promises*). While the former remains relatively objective and externalized, the latter indicates the attitude of the speaker – certainly a domain of subjectivity. Similarly, in *I think*, the judging (or attitudinal) subject is brought up. In *I am happy*, *I* alerts many possible subjectivities, but in the utterance context, emotional domains are alerted, and thus the affective subject is profiled.

Some discursive contexts represent domains of convention that lend a deictic such as *I* an extremely narrow zone of application (narrow to the momentary reality of discourse). Consider our everyday reliance on hugely schematic role representations to support the interactions we have with people who work and live in our communities; these representations concern specific capacities and not others. Further, given that each social being possesses multiple identities/identifications, if these are available to the shared reality, any of these, too, may be indexed. Jo Rubba (1996) has shown how idealized cognitive models impinge upon more “objective-world” factors in language production. Lakoff’s (1996) account of the interpretation of *I dreamt I was Brigitte Bardot and that I kissed me* involves a necessary, conventionalized partitioning of self into “(thinking,

---

<sup>11</sup> Finegan (1995) writes that it is precisely **because subjectivity is the “expression of self and the representation of a speaker’s ... perspective or point of view in discourse – what has been called a speaker’s imprint...” that it has a multitude of meanings and instantiations. Subjectivity itself is not so limited or systematized as grammatical subjecthood, but “central to emerging views of discourse – to the intersection of language structure and language use in the expression of self”.**

### *Intersubjectivity of ‘Self’ and ‘Other’ Tokens*

judging) subject” and “(corporeal) self”. Thus, despite its kaleidoscopic nature, *as we understand it* subjectivity has been conventionalized and typed in functionally, socially, significant ways.

Finally, perhaps the most convincing evidence comes from the role that conversational implicature plays in semantic change: we would not be able to infer, much less consistently lexicalize, particular senses or attitudes (aspects of subjectivity) into a lexical meaning *unless* we were able to abstract (and quite effortlessly) from the infinite variety of specific cognitive states that could plausibly accompany a previously “objective” utterance. In sum, then, it is highly plausible that specific domains of subjectivity will be indexed, or alerted, by an *I*. Thus mentions of the pronoun are nothing other than an invitation to *consider* the subjectivity associated with the pronoun (1<sup>st</sup> or 2<sup>nd</sup> person, either interlocutor) in association with the proposition offered in the surrounding utterance.

Returning to the contentious nature of the exchange in the example, we are now in a position to describe certain identitarian risks in interaction. If, as in encyclopedic cognitive semantics, the speaking subject, S, inheres in every element’s semantic value, then it can only be true that we, as interlocutors, in playing a part in the conceptualizations brought up, are always both subjects and objects of conception. This is most evident in the case of *I* and *you*. When I say *I was attacked by a dog last night*, I place myself on-stage, as the experiencer of a dog attack. To my addressee, I may therefore be objectified – observed at a notional distance, highly available as an object of contemplation. When our selves become of concern – which is particularly the case in uncomfortable or marked social situations (rather than unmarked ones; see Goffman 1959) such as insult exchanges – then we begin to play games of representation. The utterance context surrounding the use of *I* informs the subjectivity thus involved and goes “on record”, whether it is an attitude that can go on record without deep inferencing, or a short-lived statement of attitude that quickly fades. Whatever the “private” stabilities of the identities involved in the interlocution, both participants must fend with the existence of an eminently alienable representation of themselves that resides in the shared, public sphere.

To play out those possible risks and how they emerge, we address the possibility of *implicit* perspectivization, which underlies much of what can be sneakily harmful in discourse with regard to identity, and hence is relevant here. Spaces or elements representing *you*, *I*, *you think I*, and so on, are implicit in any conversational context, but particularly an adversarial one. In orthodox MST, *you think* makes Y’s R space the focus space. However, it is my contention that while there are biasing constraints, still any of these spaces can serve as a focus space, whether by explicit marking or not. Though Sanders and Spooren (1997) claim that the egocentric perspective is the default unless shifting is invited explicitly, much of this can be done implicitly.

Awareness of intersubjectivity is essential to the ability to perspectivize, and in fact, it is one condition obliging us to afford perspectivization to utterances that may not be explicitly marked. Once we abstract away from a physicalist model of

perspective (of the type that seems to underlie Langacker's analyses of objectivity and subjectivity), other phenomena, such as affective frames, and value judgments, and different instantiations of the conceptualized "object" come into play. I have already recounted Langacker's claim that every conceptualized element has at least a minimal schematic representation of a conceptualizer included within it. In some sense, this is the foundation for the awareness of the intersubjective nature of meaning, however below the level of outright consciousness this may lie.<sup>12</sup> Polite exchanges, and implicitly perspectivizable utterances, can be seen as involving processes of meaning that may be characterized as predominantly rhetorical rather than strictly semantic in a traditional sense.

How is this perspectivization afforded? If we follow the traces of the discourse, we could go so far as to say that, even if *I/you* are indexical, they still may nevertheless be relatively subjectified (if by the speaker) and objectified (for the hearer). For instance, a presumed relative conceptual objectivity is broken when the unsuspecting subject suddenly comes to suspect and subjectifies the *you* that was spoken to him or her. In Exchange 2, B has wisely perspectivized the labeling of him as a *numbskull*, rather than accepting it as part of the subjective objectivity space being built. Thus, B does not seek to correct a characterization of *him* that has been planted there, but rather chooses not to "see" his place in it at all – instead, the only common entity he purports to address is the meaning of *numbskull* – the real contest. Thus in this kind of direct and indirect insult, perspectivization works to flag certain spaces, with certain potentially invested contents, as the "shared", and hence "subjectively objective" reality space. Conditions of power and authority – in the form of who, a priori, gets automatic control of the intersubjective reality space, but also, no less in the form of what statements are allowed, play a crucial role in determining which spaces are available for assigning meaning.

The explanation of meaning processes in insult exchanges is not trivial because much of it happens implicitly, and hence passes under the scope of explicit linguistics. When we consider other domains in which self- and other- representations might be invoked with some kind of investment in the discourse, insult emerges as a particularly brusque, self-evident, and self-alerting genre. There are other, subtler contestations of identity – or manipulations of identity – which do not so perceptibly burst the social rules. As we can see even in the first exchange, at a certain point B realizes that his *I* cannot quite jibe with the other's *you* representations, so that one space must in effect be dissassociated into two, where it was not previously necessary. Interlocutor B in exchange 1 has perhaps realized at some point that he has played a part in a denigration of himself without knowing it (because he did not immediately think to contradict it, and so it stood). Verbal

---

<sup>12</sup> Langacker (1997) writes: "[I]n a speech event... each participant's construal of the situation includes the fact that the other also apprehends it in a certain manner; how the situation is to be viewed and portrayed therefore constitutes a major portion of what the interlocutors have to negotiate."

## *Intersubjectivity of 'Self' and 'Other' Tokens*

passive-aggression counts as another relatively subtle form. Whether the aggression is verbal or psychological, there is conceived to be a certain force working between the agents indicated by *I* and *you*. Are, in this case, the representations indexed by one's *I* and the other's *you* shared intersubjectively? Perhaps relatively more so in this case than in the insult case, provided that in the insult case there is the possibility for explicit, discernible resistance. Insult is by no means the unique holder of identitarian contests. It is the explicit relationship between meaning and interaction that distinguishes these contests.

In sum, the two simple argumentative dialogues have suggested some interesting points:

- New space generation can arise from conflict inhering in representations of the participating subjects.
- Different subjectivities (kinds of subjectivity) are being "indexed".
- The great schematicity of *I/you* does not prevent them from being associated with, or evoking, certain durative local meanings, by referring and/or modifying existing mental spaces. On the one hand: "semantics" (just an index) leaks into "meaning" (for participants, *I* "means" something in between the *I* in the discourse and the *I* at the moment of speaking); and yet the semantics remains secure and self-limiting precisely because it allows for the indexing process to account for changing representations as well.
- The participation of both interlocutors in the negotiation of meaning cannot be opted out of, unless the conversation itself is abandoned; however, within the interaction the authority of meaning-making may well be asymmetric at any given point in time. Certainly the explicit and implicit differentials of power inhering in any conversational situation make this a likelihood, albeit one that may manifest subtly.

One of the major insights owed to sociology and subsequent work in sociolinguistics is that it is by language that persons construct their identities. And if we see the *I* as the purest linguistic tool, or, better, indirect agent, of our own subject representations, taking Benveniste to heart, then we must acknowledge that this construction happens largely through these pronouns, and hence that the entity indexed by *I* is itself necessarily mobile, undetermined. The main question to answer, then, is to what degree, for each interlocutor, potentially disparate aspects of one's identity come to bear on the discourse at hand, and whether we can be satisfied in saying that a criterion of relevance is sufficient for a full explanation of the semantics of an utterance.

## **References**

- Benveniste, E. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.  
Fauconnier, G. 1994. *Mental Spaces*. Cambridge: Cambridge University Press.

Melinda Y. Chen

- Fauconnier, G. 1997. *Mappings in Thought and Language*. Cambridge: Cambridge University Press.
- Finegan, E. 1995. Subjectivity and subjectivisation: An introduction. In D. Stein and S. Wright, eds., *Subjectivity and Subjectivisation: Linguistic Perspectives*, 1-15. Cambridge: Cambridge University Press.
- Goffman, E. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor.
- Lakoff, G. 1996. Sorry, I'm not myself today. In G. Fauconnier and E. Sweetser, eds., *Spaces, Worlds, and Grammar*, 91-123. Chicago: University of Chicago Press.
- Langacker, R. W. 1991. Subjectification. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*, 315-342. Berlin: Mouton de Gruyter.
- Langacker, R. W. 1997. The contextual basis of cognitive semantics. In J. Nuyts and E. Pederson, eds., *Language and Conceptualization*. Cambridge: Cambridge University Press.
- Lyons, J. 1982. Deixis and subjectivity: Loquor, ergo sum? In R. Jarvella and W. Klein, eds., *Speech, Place and Action*, 101-124. New York: John Wiley.
- Rubba, J. 1996. Alternate grounds in the interpretation of deictic expressions. In G. Fauconnier and E. Sweetser, eds., *Spaces, Worlds, and Grammar*, 227-261. Chicago: University of Chicago Press.
- Sanders, J., and W. Spooren. 1997. Perspective, subjectivity, and modality from a cognitive linguistic point of view. In W.-A. Liebert, G. Redeker, and L. Waugh, eds., *Discourse and Perspective in Cognitive Linguistics*, 85-114. Amsterdam: John Benjamins.
- Schiffrin, D. 1993. 'Speaking for another' in sociolinguistic interviews. In D. Tannen, ed., *Framing in Discourse*. Oxford: Oxford University Press.
- Sweetser, E., and G. Fauconnier. 1996. Cognitive links and domains: Basic aspects of mental space theory. In G. Fauconnier and E. Sweetser, eds., *Spaces, Worlds, and Grammar*, 1-28. Chicago: University of Chicago Press.
- Traugott, E. C. 1986. From polysemy to internal semantic reconstruction. In V. Nikiforidou, M. VanClay, M. Niepokuj, and D. Feder, eds., *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, 539-550. Berkeley, CA: Berkeley Linguistics Society.
- Traugott, E. C. 1995. Subjectification in grammaticalisation. In D. Stein and S. Wright, eds., *Subjectivity and Subjectivisation: Linguistic Perspectives*. Cambridge: Cambridge University Press.
- Taylor, J. 1994. 'Subjective' and 'objective' readings of possessor nominals. *Cognitive Linguistics* 5(3): 201-242.

Melinda Y. Chen  
Department of Linguistics, UC Berkeley  
1203 Dwinelle Hall  
Berkeley, CA 94720-2650

melc@socrates.berkeley.edu

## Voiceless Tone Depressors in Nambya and Botswana Kalang'a\*

LAURA J. DOWNING and BRYAN GICK

ZAS, Berlin

University of British Columbia

### 0. Introduction

It is well known (Beach 1924, Hombert 1978, etc.) that voiced obstruents can act as tonal depressors, lowering the phonologically expected tone of a following vowel. For example, Hyman & Mathangwane (1998) and Mathangwane (1999) show that in Botswana Kalang'a (BK), word-final High tones normally double onto the toneless initial syllable of a following word. However, an intervening voiced obstruent blocks high tone doubling (Mathangwane 1999:173, Fig. (2)):

(1) Botswana Kalang'a High tone spread

a. /tʃi-pó tʃi-tʃó/ → [tʃi-pó tʃí-tʃó] 'your gift'

BUT depressor blocks high tone spread (depressor is underlined):

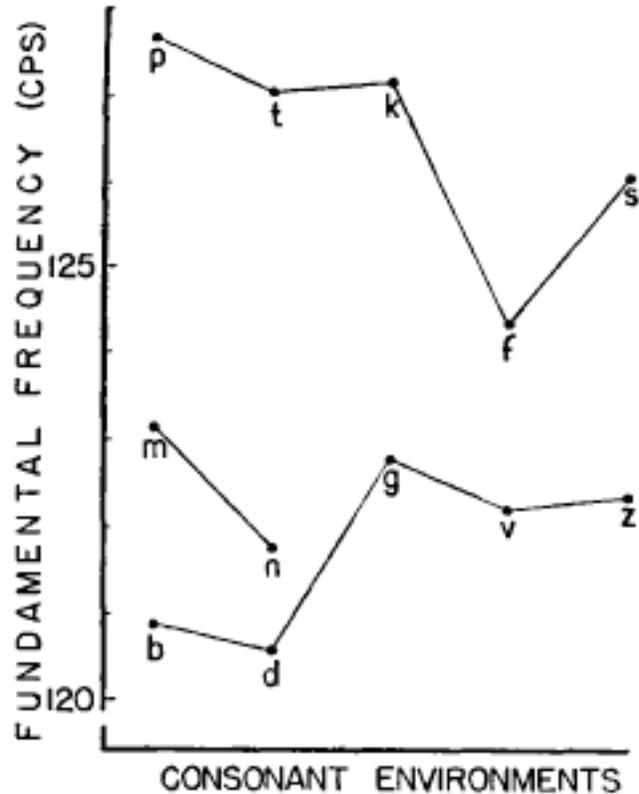
b. /z<sup>w</sup>i-pó z<sup>w</sup>i-zó/ → [z<sup>w</sup>i-pó z<sup>w</sup>i-zó] 'your (pl.) gifts'

Phonetic studies (House & Fairbanks 1953, Hombert et al. 1979) provide an explanation for tonal depression: voiced consonants lower the F<sub>0</sub> of a following vowel even in non-tonal languages like English, as shown in (2). On the basis of this phonetic interaction between tone and voice, some recent theories of phonological feature representation (Bradshaw 1999, Halle 1995) argue that voicing and low tone should be designated with a single feature.

---

\* This research was supported in part by NSF POWRE grant #SBR-9806180, and an International Research supplement to this grant, awarded to Laura Downing. The International Grant allowed her to spend May-July 2000 at the University of Zimbabwe, Harare, where the Nambya data cited in this paper was collected. Thanks to the Departments of Linguistics and of African Languages at the University of Zimbabwe, and in particular to Francis Matambirofa and Carolyn Harford, for their hospitality during this stay. Special thanks are due to Theodora Ncube, an undergraduate student at the University of Zimbabwe and native speaker of Nambya from Hwange, Zimbabwe, for her patience and insight in working on her language. Phonetic analysis was conducted at the UBC Interdisciplinary Speech Research Laboratory. Any errors of fact or interpretation in this paper are our responsibility.

- (2) Effect of consonant on F0 of following vowel  
(House & Fairbanks 1953, Fig. 2)



In this paper, we present phonetic evidence from both BK and Nambya (closely related Southern Bantu languages) that is problematic for the view that tone lowering is necessarily linked to [+voice], since it shows that both languages contain phonetically and phonologically voiceless depressors. The paper is organized as follows. In section 1, we show that contrastively voiced obstruents are tonal depressors, in both BK and Nambya, while other consonants are not. Then we show that both languages have apparently voiceless obstruents which pattern with the voiced obstruents in triggering tonal depression. In section 2, we present a phonetic study of the voiceless depressors showing first that they are phonetically voiceless, second that the voiceless depressors are phonetically distinct from the voiceless non-depressors, and third that this phonetic property is shared across both languages. Links with observations from other languages are discussed as well. Finally, in section 3, we discuss the implications of these findings for both phonological theories of tone-voice interaction and a phonetic understanding of the influence of consonantal properties on F0.

### 1. Depressor Effects in BK and Nambya

In both BK and Nambya, as in other Southern Bantu languages, voiced obstruents “interfere” with productive tone processes, lowering the tone of a following vowel

*Voiceless Tone Depressors in Nambya and Botswana Kalang'a*

in some phonological contexts.<sup>1</sup> In BK, as Hyman & Mathangwane (1998) and Mathangwane (1999) demonstrate, contrastively voiced obstruents act as High tone blockers. As shown in (3), word-final High tones productively double onto a following (toneless) syllable if the intervening consonant is voiceless (or not contrastively voiced, so sonorants and implosives pattern with voiceless obstruents for this process):<sup>2</sup>

- (3) BK tone doubling (HTS3) – not blocked by voiceless obstruents
- a. /tʃi-pó tʃi-tʃé/ → [tʃi-pó tʃí-tʃé] ‘his/her gift’
  - b. /ku- tʃá βu-sîkú/ → [ku- tʃá βú-sîkú] ‘to fear at night’
  - c. /ku- túmá βu-sîkú/ → [ku- túmá βú-sîkú] ‘to send at night’

However, as shown in (4), this tone doubling process is blocked if a contrastively voiced consonant (depressor consonant) intervenes:

- (4) BK tone doubling (HTS3) – blocked by depressor consonants
- a. /ẓ<sup>w</sup>i-pó ẓ<sup>w</sup>i-zé/ → [ẓ<sup>w</sup>i-pó ẓ<sup>w</sup>i-zé] ‘his/her gifts’
  - b. /ku- tʃá ẓ<sup>w</sup>i-pó/ → [ku- tʃá ẓ<sup>w</sup>i-pó] ‘to fear gifts’
  - c. /ku- túmá ẓ<sup>w</sup>i-pó/ → [ku- túmá ẓ<sup>w</sup>i-pó] ‘to send gifts’

A further depressor effect in BK is revealed by comparing the data in (5a) with that in (5b). As Hyman & Mathangwane show, while tone doubling is not blocked word internally by depressor consonants, the vowels following these consonants surface with a Low tone in contexts where we find High tones in comparable words that lack depressor consonants.<sup>3</sup> Note that there is only one underlying High tone in verb stems, which is arguably contrastively associated with the stem-initial vowel; other High-toned syllables within the stem are the result of tone doubling:

---

<sup>1</sup> Both Nambya and BK are linguistically and geographically neighboring Southern Bantu languages. All the Nambya data cited in this paper is from Downing (field notes). All the Botswana Kalang'a data is from Hyman & Mathangwane (1998) and Mathangwane (1999), and is based on the dialect of Joyce Mathangwane, a native speaker linguist of BK. Thanks to Joyce for helpful discussion of the BK data. See Bradshaw (1999) for a comprehensive survey of languages exhibiting depressor effects.

<sup>2</sup> Hyman & Mathangwane (1998) refer to this process as HTS3, to distinguish it from two other processes of High tone doubling (spread) which they show are not affected by the presence of voiced obstruents. In (2), (3) the representations to the left of the arrow are inputs to HTS3, not necessarily the underlying representations. In all the data cited, acute accent indicates High tone; lack of accent indicates Low tone.

<sup>3</sup> It is beyond the scope of this paper to account for why depressor consonants have different effects in different phonological domains in BK. The interested reader can consult Hyman & Mathangwane (1998) for detailed discussion. For our purposes, the point to be drawn from these data is that generally only contrastively voiced obstruents lower the tone of a following vowel.

(5) BK word medial High tone realization in High-toned verb stems

- a. Stems without depressors  
 [ku-túmá βú-sîkú] ‘to send at night’  
 [ku-fúmíká βu-sîkú] ‘to cover at night’
- b. Stems with depressors  
 (underlined; unexpected Low-toned vowels are bold)  
 [ku-dzimá βú-sîkú] ‘to extinguish at night’  
 [ku-pédza βú-sîkú] ‘to finish at night’  
 [ku-dzimúlá βu-sîkú] ‘to reduce heat at night’  
 [ku-bíganyá βu-sîkú] ‘to bury at night’  
 [ku-múlídza βu-sîkú] ‘to lift at night’

Research on Nambya tone so far reveals a similar depressor effect. As shown by the data in (6), the stem-initial High tone of a verb stem regularly doubles on to a following vowel.

(6) Nambya realization of High tone on stems without depressors

|             |                     |             |            |
|-------------|---------------------|-------------|------------|
| [ku-tálá]   | ‘to measure’        | [ku-táná]   | ‘to delay’ |
| [ku-pátúka] | ‘to burst’          | [ku-písá]   | ‘to burn’  |
| [ku-kwómá]  | ‘to dry (intrans.)’ | [ku-kángá]  | ‘to fry’   |
| [ku-fó bá]  | ‘to call’           | [ku-fótóka] | ‘to jump’  |

However, as shown by the data in (7), if the High-toned stem contains a depressor consonant, the vowel following the depressor is Low-toned in contexts where we find High tones in comparable words in (6) that lack depressor consonants:

(7) Nambya realization of High tone on stems with depressors

(underlined; unexpected Low-toned vowels are bolded)

|                       |                 |                                  |                           |
|-----------------------|-----------------|----------------------------------|---------------------------|
| [ku- <u>bo</u> bóla]  | ‘to pierce’     | [ku- <u>z</u> <sup>w</sup> imbá] | ‘to swell; to miss’       |
| [ku- <u>dz</u> elúla] | ‘to tear’       | [ku- <u>z</u> ilílika]           | ‘to sweat’                |
| [ku- <u>g</u> otfá]   | ‘to roast’      | [ku- <u>be</u> .úka]             | ‘to turn over (intrans.)’ |
| [ku- <u>dz</u> imá]   | ‘to extinguish’ | [ku- <u>y</u> uná]               | ‘to harvest; to break’    |

The data so far shows that in BK and Nambya contrastively voiced consonants trigger tonal depression. Again, as shown in (2) above, this well-known pattern is consistent with data from previous studies.

Because the correlation between voicing and pitch lowering is uncontroversially both phonetically grounded and phonologically active, several recent feature theories propose to formally encode this relationship. Peng (1992), for example, argues that the tone/voice correlation is best formalized in terms of grounding constraints (Archangeli & Pulleyblank 1994), like those in (8a). As shown in (8b), in languages where tone/voice is phonologically active, these

*Voiceless Tone Depressors in Nambya and Botswana Kalang'a*

constraints would evaluate a tautomoraic CV sequence as phonetically grounded if the consonant is [+voice] and the vowel is Low toned. However, since the representation in (8c), with the consonant [-voice] and the vowel Low toned, is ungrounded, no language should have active alternations requiring this output.

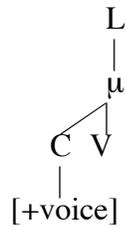
(8) Peng (1992) tone/voice grounding theory

a. (Some) Grounded Constraints:

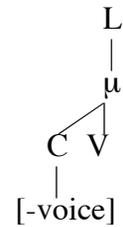
IF [+voice] THEN Low tone.

IF [-voice] THEN NOT Low tone.

b. Grounded representation



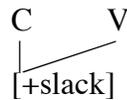
c. Ungrounded representation



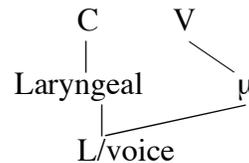
Other recent theories, Halle (1995; originally proposed in Halle & Stevens 1971) and Bradshaw (1999), formalize the correlation by defining Low tone and [+voice] as different phonetic realizations of the same phonological feature, either [+slack vocal cords] (Halle & Stevens 1971; Halle 1995) or [L/voice] (Bradshaw 1999). As shown in (9a, b), both of these theories straightforwardly predict that a contrastively voiced consonant will lower the pitch of a following vowel if [+slack] or [L/voice] is associated both with the voiced consonant and a following vowel. In contrast to Peng's grounding approach, a relation between [-voice] and Low tone is simply inexpressible in both of these theories.<sup>4</sup>

(9) Low tone = [+voice]

a. Halle (1995)



b. Bradshaw (1999)



<sup>4</sup> For Bradshaw (1999), in fact, [voice] is a privative feature, so only a correlation between [+voice] and Low tone can be formalized. For Halle (1995), voicing is not privative; [+stiff] designates both [-voice] and High tone. It is not clear to us how either of these theories accounts for the fact that [+voice] is the marked feature for voicing in these frameworks while High tone is arguably (Pulleyblank 1986) the marked feature for tone. It is beyond the scope of this paper to pursue this problem. However, see Anderson (1978) for a detailed discussion of why it is problematic to designate tone and voicing with the same feature.

All three theories make the strong prediction that only voiced consonants should lower the tone of following vowels.

It is extremely surprising, then, to find that both BK and Nambya have a set of voiceless obstruents which pattern with the contrastively voiced obstruents in acting as tonal depressors. As Mathangwane (1999) shows, in BK there is a contrast between ‘non-depressor’ aspirated voiceless stops and ‘depressor’ aspirated stops (P<sup>h</sup>, T<sup>h</sup>).<sup>5</sup> As shown in (10a), the non-depressors, as expected with voiceless obstruents, do not block HTS3 and can in general be followed by High tones. The ‘depressor’ aspirated stops, in contrast, pattern with other depressor consonants in blocking HTS3 and in being necessarily followed by a Low tone on the surface. This is shown in (10b).<sup>6</sup>

(10) Botswana Kalang’a depressor vs. non-depressor voiceless aspirates (depressors are underlined; compare the tone of the bolded vowels in the depressor vs. non-depressor sets)

a. *Non-depressor aspirates and non-aspirates*

- |       |                                          |   |                                          |                            |
|-------|------------------------------------------|---|------------------------------------------|----------------------------|
| (i)   | /ku- <u>f</u> ímá p <sup>h</sup> ilé/    | → | [ku- <u>f</u> ímá p <sup>h</sup> ílé]    | ‘to hate a bad singer’     |
| (ii)  | /ɲóká N- <u>l</u> efú/                   | → | [ɲóká ndéfú]                             | ‘long snake(s)’            |
| (iii) | /p <sup>h</sup> óko jáŋgu/               | → | [p <sup>h</sup> ókó jáŋgu]               | ‘my (castrated) billygoat’ |
| (iv)  | /-t <sup>h</sup> ánt <sup>h</sup> anula/ | → | [-t <sup>h</sup> ánt <sup>h</sup> ánúla] | ‘unstitch; unbraid’        |

b. *Depressor aspirates*

- |       |                                              |   |                                       |                        |
|-------|----------------------------------------------|---|---------------------------------------|------------------------|
| (i)   | /ku- <u>f</u> ímá <u>P</u> <sup>h</sup> ené/ | → | [ku- <u>f</u> ímá P <sup>h</sup> ené] | ‘to hate a steenbuck’  |
| (ii)  | /ɲóká N- <u>t</u> atú/                       | → | [ɲóká T <sup>h</sup> atú]             | ‘three snakes’         |
| (iii) | / <u>T</u> <sup>h</sup> éko jáŋgu/           | → | [T <sup>h</sup> éko jáŋgu]            | ‘my hiccup’            |
| (iv)  | /kú <u>T</u> <sup>h</sup> uβúla/             | → | [kúT <sup>h</sup> uβúla]              | ‘pluck off, of leaves’ |

In Nambya, there is a contrast between two ‘f’s, a ‘depressor f’ and a non-depressor.<sup>7</sup> As shown in (11a), High-toned verb stems that begin with the non-

<sup>5</sup> Mathangwane (1999) transcribes these depressor aspirates as having a breathy voiced release. We are adopting a different transcription here, for reasons to be explained in section 2.3.

<sup>6</sup> In BK, as in many Bantu languages, /l/ regularly alternates with [d] after a homorganic nasal. /N/ is the class 9/10 prefix required on adjectives and other modifiers of nouns in this class (which often themselves have this prefix). ‘Depressor’ aspirates are the synchronic as well as the diachronic output of N+voiceless stop sequences. (See section 3.1., below, for more examples and discussion.)

<sup>7</sup> The distinction between the two ‘f’s is recognized in the orthography, with the depressor written ‘fh’ to distinguish it from the non-depressor ‘f’. (An ‘h’ following a letter is a Southern Bantu convention for indicating a sound which is breathy voiced and/or a tonal depressor.) However, Theodora Ncube did not pronounce all the words the dictionary spells ‘fh’ with the depressor ‘f’, so the pronunciations here are not likely to be spelling pronunciations. In fact, it was not clear to me how widely known Nambya orthography is among native speakers, since Theodora felt that ‘to sew’ and ‘to become rich’, for example, are to be spelled and pronounced identically, even though the dictionary spells them differently and there is a clear tonal difference in the pronunciation. It is also not clear to me whether the differences between Theodora’s pronunciation and that suggested

## *Voiceless Tone Depressors in Nambya and Botswana Kalang'a*

depressor have a High tone on the first two syllables (cf. (6)). In contrast, stems that begin with 'depressor f' (F) pattern with the other depressor consonant-initial stems (cf. (7)) in having a Low tone on the first syllable and a High tone on the second, as shown in (11b).

(11) Nambya, realization of High tone in stems with non-depressor 'f' vs. depressor 'F'

a. *Non-depressor 'f'*

|                     |                  |                      |                                  |
|---------------------|------------------|----------------------|----------------------------------|
| [ku- <u>f</u> úmá]  | 'to become rich' | [ku- <u>f</u> úpísa] | 'to shorten'                     |
| [ku- <u>f</u> úndá] | 'to learn'       | [ku- <u>f</u> úlá]   | 'to blow s.t. out of s.o.'s eye' |
| [ku- <u>f</u> á]    | 'to die'         | [ku- <u>f</u> ílá]   | 'to die for'                     |

b. *Depressor 'F' (underlined)*

|                      |             |                    |                 |
|----------------------|-------------|--------------------|-----------------|
| [ku- <u>F</u> umá]   | 'to sew'    | [ku- <u>F</u> ilá] | 'to spit'       |
| [ku- <u>F</u> ulíla] | 'to thatch' | [ku- <u>F</u> ulá] | 'to work metal' |

The voiceless depressors raise research questions which we pursued by undertaking a phonetic analysis of recordings of the Nambya data in (11). First, are the 'voiceless' depressors in (11b) actually phonetically voiceless or does some measurable degree of (breathy) voicing distinguish them from the non-depressors in (11a)? If it does turn out that the voiceless depressors are truly voiceless, is there another phonetic property which distinguishes the depressors and non-depressors? Finally, do the voiceless depressors of Nambya share any common phonetic property with the voiceless depressors of BK? The results of our phonetic study are presented in the next section.

## **2. Phonetics of Voiceless Depressors in Nambya and BK**

An acoustic analysis was conducted to test for phonetic differences between the two types of 'f' (depressor and non-depressor) in Nambya. In section 2.3, these results are compared to findings from Mathangwane's (1999) study of BK aspirated stops.

### **2.1. Methods**

Field recordings were collected on standard cassette tapes from a single female native speaker of Nambya, who produced a variety of words in isolation. Recordings were digitized at 44.1kHz onto an Apple PowerBook G3 using SoundEdit™ 16 v2 audio editing software. 16 tokens of /f/ and 13 tokens of /F/ were chosen from a single recording session. In all examples the target segment occurs word-internally, in the context #Cu\_uCV..., with the first syllable of the word unstressed. Acoustic analysis of these tokens was conducted using Scicon's

---

by the dictionary spelling reveals a dialect difference or a difference in the pronunciation of younger compared to older speakers.

MacQuirer v4.9.7. Standard two-way analyses of variance (ANOVAs) calculated using StatView v5.0 were used to compare the two types of 'f'.

As was evident both from the spectrograms and from FFTs calculated on the midpoint of frication, two distinct spectral peaks were identified for both types of 'f': One longer, more prominent band with a maximum amplitude averaging around 5.5kHz, and the other with a maximum amplitude averaging around 14.5kHz. Since this two-peak pattern is typical of [f], and since the very high frequency band may be perceptually relevant in distinguishing fricatives (Tabain 1998), both were analyzed in this study. Thus, nine different measurements were taken to compare the two types of 'f':

- Spectral (FFT) measures:
  1. Bandwidth (Hz) of high freq. band at midpoint of frication (cf. Tabain 1998)
  2. Bandwidth (Hz) of low frequency band at midpoint of frication
  3. Frequency (Hz) at peak of high frequency band at midpoint of frication
  4. Frequency (Hz) at peak of low frequency band at midpoint of frication
  5. Peak amplitude (dB) of high frequency band at midpoint of frication
  6. Peak amplitude (dB) of low frequency band at midpoint of frication
  
- Temporal measures:
  7. Duration (ms) of high frequency band
  8. Duration (ms) of low frequency band
  9. Total duration (ms) of frication

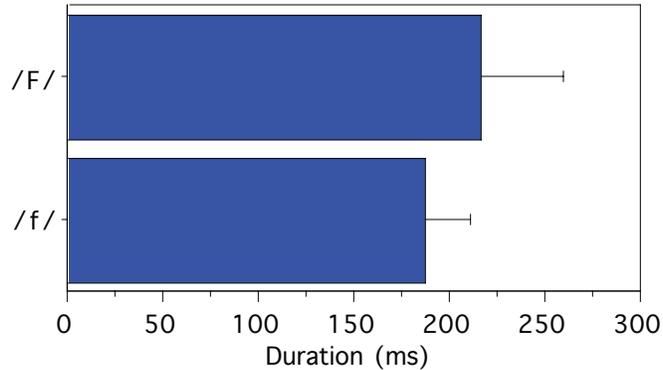
## 2.2. Results

First, no difference was found between Nambya /f/ and /F/ in terms of periodicity during frication; except for miniscule sporadic spillover of voicing from the preceding vowel into the onset of frication (up to three periods) appearing in both types, both types of 'f' were completely voiceless throughout.

According to ANOVA results, both types of 'f' were found to be the same for all six of the spectral measures. However, the total duration of frication noise (and, proportionally, the durations of each of the two component bands) was found to be significantly greater ( $p < .05$ ) for depressor /F/ than for non-depressor /f/ (mean 217ms vs. 187ms, respectively, as shown in (12)).

*Voiceless Tone Depressors in Nambya and Botswana Kalang'a*

- (12) Mean duration of frication noise in Nambya depressor /F/ vs. non-depressor /f/ (error bars = standard deviation)



**2.3. Comparison of Nambya Results with BK**

The above results show that there is indeed a phonetic difference between Nambya depressor /F/ and non-depressor /f/, with the depressor showing significantly greater frication duration than the non-depressor. In this section, we compare these results to the aspirated series reported for BK by Mathangwane (1999).

First, Mathangwane refers to the tone-depressor aspirated stops /P<sup>h</sup>, T<sup>h</sup>/ in BK as the “breathy aspirates”, identifying them as being phonetically distinct from the non-depressor aspirate series. It is important to note that she uses the term “breathy” here “for ease of identification,” and does not claim that these stops are actually voiced or breathy-voiced stops (she has since further verified their voiceless status in personal communication). She does, however, describe the differences between the two types of aspirated stop in some detail. The main difference she cites (aside from the pitch effect on the following vowel) is that the tone-depressor aspirates “have a longer duration of noise in the high frequencies than the regular aspirates after the vowel onset...,” citing a difference of 32 ms between the mean durations. This difference in duration is almost identical to the 30-ms difference between frication duration in the two types of ‘f’ observed above for Nambya.

Unfortunately, as the data from Mathangwane’s study is no longer available, we are unable to support our interpretation of her results. In particular, questions remain as to what is meant by “noise in the high frequencies,” and “after the vowel onset.” Both of these will have to be addressed (perhaps using new data) before the connection between the Nambya and BK stops can be finally confirmed. However, we believe that the existence of this phonetic connection linking the frication duration of the Nambya f’s and of the BK aspirate release bursts further supports the view that voicelessness and tone depression may be compatible for certain classes of voiceless segments.

This possibility is further supported by F0 lowering effects found in previous phonetic studies of aspiration and frication. While unaspirated voiceless stops

have generally been viewed as uncontroversially raising vowel F0 relative to voiced stops, the results for aspirated stops and fricatives have been much less clear.

Zee (1980:90) cites a number of cases where, contrary to the typical pattern, aspirated stops have been observed to have a lowering effect on the F0 of the following vowel, as compared with the effect of the unaspirated equivalent. In a study of Thai stops (Erickson 1975), while 8 of 11 subjects showed a higher F0 onset for vowels following an aspirated stop than for those following an unaspirated stop, the remaining three subjects showed the opposite effect. This F0 lowering effect in Thai was also observed by Gandour (1974). Kagaya & Hirose (1975) observed a similar F0 lowering effect following aspirated stops for a speaker of Hindi. Particularly interesting for the case of aspirated stops is that of Madurese (Cohn 1993, Cohn & Lockwood 1994). Cohn & Lockwood (1994) measured both closure duration and VOT of nasal, voiced, voiceless unaspirated, and voiceless aspirated stops, as well as F0 of following vowels, of two speakers of Madurese. Their study found that aspirated stops not only lowered the F0 of the following vowel (by 10Hz for their first subject and 40Hz for the second), but that the aspirated stops were 23-40ms longer in total duration (closure plus VOT) than the voiced and voiceless unaspirated stops, respectively. This finding further supports the connection between duration of voiceless aspirates/fricatives and F0 lowering.

Likewise for /f/, House & Fairbanks (1953) observed that the F0 of vowels following voiceless consonants was uniformly significantly higher than those following voiced consonants, with the single exception of /f/, which grouped statistically with /m/ and /g/ in its effect on following vowel F0 (see graph in (1) above). Beyond this, other studies in this literature tended not to include /f/, leaving little comparative data.

These studies lend additional support to the notion that distinct mechanisms are employed across different languages, and even different speakers of the same language, for producing aspiration/frication, and that these production mechanisms affect the F0 of following vowels in distinct ways, which in turn may become phonologized in some tone systems.

### **3. Implications for the Phonology and Phonetics of Tonal Depression**

#### **3.1. Problems with Adapting Current Theories of Tone-Voice Interaction**

These phonetic results are clearly problematic for the phonological theories illustrated in (8) and (9) which tie Low tone to [+voice]. While it has been observed before (Zulu (Traill et al. 1998), Musey (Shryock 1995), Swati (Schachter 1976)) that voiceless segments can trigger depressor effects on a following vowel, the BK and Nambya voiceless depressors are fundamentally different from these cases. In the Nguni languages (Zulu and Swati) and in Musey (Shryock 1995), pitch lowering is motivated diachronically, as the synchronically voiceless consonants triggering tonal depression were historically voiced. Even though the voicing has been lost on the consonant, it is, in a sense, preserved in

*Voiceless Tone Depressors in Nambya and Botswana Kalang'a*

the tone of the following vowel, as we might predict could happen if voice and Low tone are defined by a single feature. As shown in (13), it is plausible to propose the historically voiced consonants are also synchronically voiced in the input, even though, in the output, the L/voice feature is associated only with the following vowel, since this simply involves reassociation of a single input feature ([L/voice], as shown here, or [+slack]):

(13)

- a. Input (historically motivated)      b. Output (synchronically accurate)



As Bradshaw argues, further evidence that Nguni depressor consonants are best represented as synchronically voiced in the input comes from that fact that they productively alternate with phonetically voiced consonants for other phonological processes like labial palatalization.<sup>8</sup> As shown in (14), the depressor labial stop of Swati (phonetically voiceless, unaspirated [p]) alternates with the phonetically voiced affricate [dʒ] in a palatalizing context:

- (14) Swati palatalization: [p] ~ [dʒ] (Bradshaw 1999, p. 158, fig. (20a)); [p] is a depressor)

[sigupu]      ‘calabash’  
 [sigudʒini]      ‘in the calabash’  
 [sigudʒana]      ‘little calabash’

If the depressor [p] were voiceless in the input, it would be difficult to explain why it should acquire voicing in this context. For reasons like these, Bradshaw argues that the strong correlation between voicing and pitch lowering predicted by having a single Low/voice feature is well motivated at an abstract level in languages like Swati, even though it is contradicted by the surface phonetics.

However, it is not plausible to argue that the phonetically voiceless depressors of BK and Nambya are phonologically voiced. As shown in (15), the voiceless depressors of Nambya and BK are historically as well as synchronically voiceless.<sup>9</sup>

<sup>8</sup> See Downing (1999) and Sibanda (1999) and references cited in both these works for analyses and further discussion of labial palatalization in Nguni languages.

<sup>9</sup> The BK data is from Mathangwane (1999; p. 162, Fig. (32)); the Nambya data is from Downing (field notes). The source of the Proto-Bantu cognates is the Tervuren Proto-Bantu database (Coupez, et al. 1998).

(15) Proto-Bantu cognates (u is the Proto-Bantu close high back vowel)

| <u>Proto-Bantu</u> | <u>Botswana Kalang'a</u> | <u>Nambya</u> | <u>Gloss</u>      |
|--------------------|--------------------------|---------------|-------------------|
| -túma              | -t <sup>h</sup> úmá      | -Fumá         | 'sew'             |
| -túija; -tua-      | -t <sup>hw</sup> á       | -Filá         | 'to spit'         |
| -túda              | -t <sup>h</sup> úlá      | -Fulá         | 'to work metal'   |
| -tonta             | -doT <sup>h</sup> a      | -dona         | 'drip'            |
| -tanta ('cross')   | -taT <sup>h</sup> a      | -tana         | 'climb'           |
| -póbida            | -P <sup>h</sup> obéla    | ?             | 'sink, as in mud' |

In Nambya, the source of the 'depressor f' in Proto-Bantu is a voiceless coronal stop followed by the close high back vowel. In BK, as Mathangwane (1999) shows, the source of the depressor voiceless aspirates is (typically) a nasal-voiceless stop sequence. Further, there is no evidence that the voiceless depressor consonants pattern with voiced consonants in the synchronic phonology. On the contrary, in BK there are also a few synchronic alternations showing that voiceless aspirates are the surface output of an input nasal-voiceless stop sequence. As we saw in (10b), above, the class 9/10 agreement prefix is /N-/ (a nasal that surfaces homorganic with a following consonant). Adjectives and other stems that begin with voiceless stops when preceded by other class prefixes regularly surface with a voiceless aspirated depressor in class 9/10.<sup>10</sup> More examples of this are given in (16):

(16) BK alternations between voiceless stop and depressor aspirate in Class 9/10 (alternating segments are underlined for ease of comparison)

- |                                            |                                                   |
|--------------------------------------------|---------------------------------------------------|
| a. /ku-ténga/ 'to buy'                     | /T <sup>h</sup> éngo/ 'purchase price (cl. 9)'    |
| b. /βa-tatú/ 'three (cl. 2)'               | /T <sup>h</sup> atú/ 'three (cl. 9/10)'           |
| c. /tʃi-paβí/ 'broad; wide (cl. 7)'        | /P <sup>h</sup> aβí/ 'broad; wide (cl. 9/10)'     |
| d. /m-púzu/ 'wild sour raisin tree (cl.3)' | /P <sup>h</sup> úzu/ 'sour wild raisins (cl. 10)' |

There is, then, no phonological motivation independent of tonal depression for proposing that the phonetically voiceless depressors of BK and Nambya are phonologically voiced.

Another alternative analysis that would also allow the L/voice correlation to be maintained is to propose that the vowels following the voiceless depressors of BK and Nambya are lexically associated with a Low tone. Bradshaw argues persuasively that this approach is necessary to explain some grammatical depressor effects in Swati described by Rycroft (1980). As illustrated by the data in (17), in constructions like the imperative, the (long) penult vowel of a toneless verb stem must have a rising tone (this is the usual realization of a High tone on a penult vowel preceded by a depressor consonant), no matter what consonant

<sup>10</sup> The nasal class 9/10 prefix does not surface before voiceless consonants. However, it does surface in other contexts: m-bili 'two (cl. 10)', n-dedu 'beard (cl.9)', etc.

*Voiceless Tone Depressors in Nambya and Botswana Kalang'a*

precedes the vowel. (The consonant beginning the penult syllable in all these cases is both phonologically and phonetically voiceless):<sup>11</sup>

- (17) Swati grammatical depression (unexpectedly ‘depressed’ penult is bolded) (Bradshaw 1999, p 100, Figs. (41), from Rycroft (1980))

[kho.tsàá.ma] ‘bend down!’  
 [pha.phàá.ma] ‘wake up!’  
 [tʃa.phàá.ta] ‘mock!’

Since pitch lowering here is predicted by the morphological context, not the phonological context, it can best be accounted for by stipulating that a Low tone is lexically associated with the penult in this construction.

This approach could also be made to work for Nambya and BK. As shown for the Nambya examples in (18), if a vowel preceded by a voiceless depressor consonant is lexically associated with a Low tone, it will correctly block association of a stem High tone to that syllable:

- (18) Nambya, lexical Low tone on vowel accounts for ‘depressor f’ vs. ‘non-depressor f’

a. Lexical Low tone

vs.

b. No lexical Low tone

H  
 / \  
 -F u m a    ‘to sew’  
 |  
 L

H  
 / \  
 -f u m a    ‘to become rich’

However, lexically associating a Low tone with these vowels fails to explain why the lexical Low tone only occurs with particular consonants. More importantly, it fails to explain why these consonants share a phonetic property.<sup>12</sup>

<sup>11</sup> As Rycroft (1980) notes, the rising tone on the penult is only found on toneless verb stems with no depressor consonant elsewhere in the stem. See Bradshaw (1999) for an analysis of this restriction. While it is worth noting that a study by Traill (1990) shows that the Swati “depressorless” rising tones are phonetically distinct from the rising tones following depressor consonants, the results of this study do not clearly affect the phonological analysis sketched here.

<sup>12</sup> A further problem with this approach is that it is unclear how the stem High tone could spread beyond the lexically associated Low tone in BK words like those cited in (5b), e.g.,

bíganyá.  
 | |  
 H L

### 3.2. A New Phonetic Correlate of Lowered F0?

Based on the results of this study, as well as those of previous studies such as the ones cited in section 2.3 above, we believe that there exists evidence of at least two distinct mechanisms for producing voiceless frication, one of which produces the well-known pitch-raising effects on a following vowel, and the other of which has a lowering effect. We believe that it is this latter effect that is responsible for the tone depressor effects in Nambya and BK.

The exact properties of these mechanisms remain to be elucidated in future work.

## 4. Conclusion

In this paper we have shown that both BK and Nambya contain a set of consonants which are phonetically voiceless yet have a depressor effect on the pitch of the following vowel. This finding is problematic for phonological theories which posit that only contrastively voiced segments can induce tone lowering, since the depressor segments are neither historically nor synchronically voiced. It is also phonetically unexpected, since most phonetic studies of the effect of consonants on F0 show that voiceless consonants raise rather than lower the pitch of the following vowel. However, we have also shown that these results correspond with conflicting results in the phonetics literature with respect to vowel F0 specifically following voiceless aspirated stops and the voiceless fricative /f/; we hope that this study will help to clarify these phonetic issues. While more phonetic work on these languages clearly needs to be done, this study suggests that voicing is not the only consonantal property which can lower the pitch of a following vowel.

## References

- Anderson, Stephen. 1978. Tone features. In Victoria A. Fromkin, ed., 133-175.
- Archangeli, Diana and Douglas Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: The MIT Press.
- Beach, D.M. 1924. The science of tonetics and its application to Bantu languages. *Bantu Studies* 2: 75-106.
- Bradshaw, Mary. 1999. A crosslinguistic study of consonant-tone interaction. Ph.D. dissertation, The Ohio State University.
- Cohn, A.C. 1993. Voicing and vowel height in Madurese: A preliminary report. In J. Edmondson and K. Gregerson, eds., *Tonality in Austronesian Languages*, 107-121. Oceanic Linguistics Special Publication No. 24. Honolulu: University of Hawaii Press.
- Cohn, A.C. and K. Lockwood. 1994. A phonetic description of Madurese and its phonological implications. *Working Papers of the Cornell Phonetics Laboratory* 9: 67-92.

*Voiceless Tone Depressors in Nambya and Botswana Kalang'a*

- Coupez, A., Y. Bastin, and E. Mumba. 1998. Reconstructions lexicales bantoues 2 / Bantu lexical reconstructions 2. Filemaker Pro database. Tervuren: Musée royal de l'Afrique centrale.
- Downing, Laura J. 1999. Prosodic stem  $\neq$  prosodic word in Bantu. In T. Alan Hall and Ursula Kleinhenz, eds., *Studies on the Phonological Word*, 73-98. Amsterdam: John Benjamins.
- Erickson, D. 1975. Phonetic implications for a historical account of tonogenesis in Thai. In J.G. Harris and J.R. Chamberlain, eds., *Studies in Tai Linguistics in Honor of W. J. Gedney*, 100-111. Bangkok, Central Institute of English Language Office of State Universities.
- Fromkin, Victoria A., ed. 1978. *Tone: A Linguistic Survey*. New York: Academic Press.
- Gandour, J.T. 1974. Consonant types and tone in Siamese. *Journal of Phonetics* 2: 337-350.
- Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26: 1-46.
- Halle, Morris and K. Stevens. 1971. A note on laryngeal features. *RLE [MIT] Quarterly Progress Report* 101: 198-312.
- Hombert, Jean-Marie. 1978. Consonant types, vowel quality and tone. In Victoria A. Fromkin, ed., 77-111.
- Hombert, Jean-Marie, John Ohala, and W. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55: 37-58.
- House, A. and G. Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *JASA* 25: 105-113.
- Hyman, Larry M. and Joyce T. Mathangwane. 1998. Tonal domains and depressor consonants in Ikalanga. In L.M. Hyman and C.K. Kisseberth, eds., *Theoretical Aspects of Bantu Tone*, 195-230. Stanford: CSLI.
- Kagaya, R. and H. Hirose. 1975. Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatics* 9: 27-46.
- Mathangwane, Joyce T. 1999. *Ikalanga Phonetics and Phonology: A Synchronic and Diachronic Study*. Stanford: CSLI.
- Peng, L. 1992. A unified theory of tone-voice. Ph.D. dissertation, University of Arizona.
- Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: Reidel.
- Rycroft, David. 1980. The Depression feature in Nguni languages and its interaction with tone. Grahamstown, Rhodes University, Communication No. 8, Department of African Languages.
- Schachter, Paul. 1976. An unnatural class of consonants in siSwati. *SAL*, suppl. 6: 211-220.
- Shryock, A. 1995. Investigating laryngeal contrasts: An acoustic study of the consonants of Musey. Ph.D. dissertation, UCLA.
- Sibanda, Galen. 1999. Palatalization of labials in the Ndebele passive. Ms., University of California, Berkeley.

Laura J. Downing and Bryan Gick

- Tabain, M. 1998. Non-sibilant fricatives in English: Spectral information above 10kHz. *Phonetica* 55: 107-130.
- Trail, A. 1990. Depression without depressors. *South African Journal of African Languages* 10: 166-172.
- Trail, A., J. Khumalo, and P. Fridjhon. 1987. Depressing facts about Zulu. *African Studies* 46: 255-274.
- Zee, E. 1980. The effect of aspiration on the F0 of the following vowel in Cantonese. *UCLA Working Papers in Phonetics* 49: 90-97.

Laura J. Downing  
Zentrum für Allgemeine Sprachwissenschaft,  
Typologie und Universalienforschung (ZAS)  
Jägerstraße 10-11  
D-10117 Berlin  
Germany

[downing@zas.gwz-berlin.de](mailto:downing@zas.gwz-berlin.de)

Bryan Gick  
Department of Linguistics  
1866 Main Mall, Buchanan E270  
University of British Columbia  
Vancouver, British Columbia V6T 1Z1  
Canada

[gick@interchange.ubc.ca](mailto:gick@interchange.ubc.ca)

## Object Marking and Agentivity in Navajo Causatives\*

SUZANNE GESSNER  
*University of British Columbia*

### 0. Introduction

The syntactic and semantic analysis of complex predicates is a topic that has received much attention in recent literature (e.g. Alsina, Bresnan, and Sells 1997). This paper aims to contribute to the debate by examining the relation between argument structure and morphology in one type of complex predicate in Navajo: the causative construction.<sup>1</sup>

Morphological causatives in Navajo fall into two types:<sup>2</sup>

- a) unaccusative verbs, which are causativized by adding the *ʔ*- classifier prefix; and
- b) unergative verbs, which, in addition to *ʔ*-, also require a *y*- prefix, as well as an object marker representing the “causee”.

We will mainly focus on the unergative verbs because the object marking in causativized unergatives is unusual: the 3rd person prefix *bi*- is used with 1st/2nd person subjects, rather than the expected zero marking, and *bi*- also appears with 3rd person subjects, instead of the expected *yi*-. We will see how these peculiar facts fall out from the general principles of object marking in Navajo, rather than being peculiar to the causative construction. In addition, this analysis lends support to the notion of argument sharing in complex predicates.

The remainder of section 0 presents relevant background information on the verbal morphology of Navajo. Section 1 presents the causative data, with previous analyses discussed in section 2. The analysis will be outlined in section 3, and section 4 summarizes the findings.

---

\* Thanks to Henry Davis and Gunnar Ólafur Hansson for their valuable input, and audiences at UBC and BLS 27 for their comments. This research is supported by Doctoral Fellowship 752-2000-2102 from the Social Sciences and Humanities Research Council of Canada and a Killam Trust Predoctoral Fellowship.

<sup>1</sup>Navajo is a language of the Apachean subgroup of Athapaskan, spoken in the southwestern United States.

<sup>2</sup>Navajo also exhibits syntactic causatives which are formed on transitive verbs, but these types of causatives are beyond the scope of this paper.

## 0.2. Navajo verb morphology

Athapaskan morphology is heavily prefixing. Verbs consist of a stem (root plus suffix, if any) that is generally monosyllabic, to which several prefixes may be added. Traditionally, in the Athapaskan literature, verbs have been represented using a template model.<sup>3</sup> The template for Navajo is shown in (1).

(1) Navajo verb template (Young and Morgan 1992)

|          |              |      |
|----------|--------------|------|
| disjunct | conjunct     | stem |
| prefixes | prefixes     |      |
| 0-1-2-3- | 4-5-6-7-8-9- | stem |

0: pronominal (object of a postposition or the possessor of a verb-prefixed noun)

1: postpositional, adverbial-thematic, nominal; reflexive; reversionary; semeliterative

2: iterative

3: distributive plural

**4: direct object pronouns**

5: subject pronouns (only 3rd person impersonal, spatial or indefinite)

6: thematic and adverbial elements

7: modal-conjugation markers

8: subject pronouns (1st/2nd/3rd persons singular and duoplural)

**9: classifiers (voice/valence markers)**

Of the many prefixes, the two that will concern us the most are the *bi-* 3rd person object prefix (in position 4) and the *#*- classifier or valency prefix (in position 9).<sup>4</sup> Word order in Navajo is SOV, but subject and object NPs are often absent, with the verb alone forming a grammatical sentence.

### 1. Causative data

In Navajo, causatives and all other transitivized intransitive verbs are formed by the addition of the transitivizing *#*- classifier prefix. Causatives are sensitive to

---

<sup>3</sup>More recently, non-templatic models of the verb have been proposed (see McDonough 1990, Hale 1997 and Rice 2000a).

<sup>4</sup>The term “classifier” is a misnomer in that the classifiers do not perform a classificatory function; rather, they mark voice and valency.

*Object Marking and Agentivity in Navajo Causatives*

the unaccusative (subject≠agent)/unergative (subject=agent) distinction.<sup>5</sup> Examples (2)-(5) illustrate causativization of unaccusative intransitives.<sup>6</sup>

- (2) a. Tʉ'óóʉ k'í-ní-dláád. (Hale and Platero 1996:4)  
 rope k'í-NPF:3-break:PERF  
 'The rope broke.'
- b. Tʉ'óóʉ k'í-i-ní-ʉ-dláád. (Hale and Platero 1996:4)  
 rope k'í-3-NPF:3-ʉ-break:PERF  
 'He broke the rope.'
- (3) a. Tóshjeeh si-ts'il. (Hale 1997:50)  
 barrel SPF:3-shatter:PERF  
 'The barrel shattered, broke to pieces.'
- b. ʉeets'aa' sé-ʉ-ts'il. (Hale 1997:50)  
 dish SPF:3:1s-ʉ-shatter:PERF  
 'I shattered the dish.'
- (4) a. Tin yí-yíəđjəđ'. (Hale 1997:50)  
 ice YPF:3-melt:PERF (<-ghíəđjəđ')  
 'The ice melted.'
- b. Yas yí-ʉ-híəđjəđ'(<-ghíəđjəđ') (Hale 1997:50)  
 snow 3:YPF:1s-ʉ-melt:PERF  
 'I melted the snow.'
- (5) a. Kō' n-eez-tsiz. (Hale and Platero 1996:3)  
 fire n-SPF:3-extinguish:PERF  
 'The fire went out.'
- b. Kō' n-é-ʉ-tsiz. (Hale and Platero 1996:4)  
 fire 3:n-SPF:1s-ʉ-extinguish:PERF  
 'I put the fire out.'

Forming a causative from unergative verbs likewise involves addition of the *ʉ*-classifier. However, in addition to the classifier, a prefix *y*- and a set of object markers representing the "causee" are also required. This is shown in the following pairs of intransitive-causative sentences in (6)-(7).

<sup>5</sup>Rice (2000b) points out that while analyses of Slave (Rice 1991) and Navajo (Hale and Platero 1996) have argued that the causative construction with the *ʉ*-classifier provides a diagnostic for unaccusativity/unergativity, the same is not true of all Athapaskan languages. She shows that in Ahtna, both unergative and unaccusative verbs can enter into this construction.

<sup>6</sup>Although Hale and Platero (1996) and Hale (1997) do not provide a list of abbreviations used, we are assuming the following: NPF= *ni*- perfective prefix, PERF= perfective stem, SPF= *si*-perfective prefix, YPF= *yi*-(*ghi*-) perfective prefix, IMP= imperfective (zero-marked) prefix, CI= continuative imperfective stem, and PROG= progressive prefix or stem.

- (6) a. 'Awéé' naa-ghá (Hale and Platero 1996:4)  
 baby na-IMP:3-walk:sg:CI  
 'The baby is walking around.'
- b. 'Awéé' na-b-ii-sh-**u**-á (Hale and Platero 1996:4)  
 baby na-**3-y**-IMP:1s-**u**-walk:sg:CI  
 'I am walking the baby around (i.e. making it walk).'
- (7) a. 'Awéé' d-ee-za' (Hale and Platero 1996:4)  
 baby d-SPF:3-belch:PERF  
 'The baby burped.'
- b. 'Awéé' bi-di-y-é-sa' (<...-**u**-za')<sup>7</sup> (Hale and Platero 1996:5)  
 baby **3-d-y**-SPF:1s-**u**-belch:PERF  
 'I burped the baby.'

Addition of the classifier prefix alone is not enough to achieve causativization, as shown by the ungrammatical example in (8b).

- (8) a. 'Awéé' yi-dloh. (Hale 1997:53)  
 baby PROG:3-d:laugh:PROG  
 'The baby is laughing.'
- b. \*(Shí) 'awéé' yishdloh. (< gh-sh-**u**-dloh) (Hale 1997:53)  
 \*'I laugh the baby.' (i.e. 'I make the baby laugh.')
- c. (Shí) 'awéé' biyeeshdloh. (< **bi-y**-gh-sh-**u**-dloh) (Hale 1997:53)  
 (I) baby **3-y**-PROG:1s-**u**-d:laugh:PROG  
 'I make the baby laugh.'

The object marking found in the unergative causative construction differs from ordinary 3rd person object marking in that the overt *bi-* prefix is used with 1st or 2nd person subjects, rather than the expected zero marking, and *bi-* also appears with 3rd person subjects, instead of the expected *yi-*.

## 2. Previous analyses

Hale and Platero (1996) call the *y-* prefix a "causative" morpheme and suggest that *bi-* object prefix is present as if it were attached to an incorporated postposition. Hale (1997) expands on this idea and draws on Case Theory (Bittner and Hale 1996) to argue that the *y-* prefix seen in causativized unergatives is itself an incorporated postposition, with the *bi-* prefix serving as its object. In other words, the *bi-* prefix is a position 0 prefix rather than a position 4 direct object marker prefix. This can be seen more clearly in the verb template in (9), repeated from (1) above, with the relevant morpheme positions shown in bold.

<sup>7</sup>Note that the classifier /**u**/ is not always visible in the surface form due to phonological interactions with the verb stem.



- (12) *biníʔdaah* ‘you are seating him’  
 cf. *nídaah* ‘you are in the act of sitting down’  
 (Young and Morgan 1987:65)

In this example, both the *#*-classifier and *bi*-object prefix are present but a *y*-prefix is not discernible.

Assuming that the *y*-prefix is present, one possibility is that it is a “peg element” (a type of epenthetic element), which frequently appears in the Navajo verb because syllable structure requirements. More likely, however, is the hypothesis that this prefix is a thematic element required by these particular verb stems in the causative construction. Such thematic prefixes obligatorily occur in the causatives of other Athapaskan languages (such as Slave syntactic causatives; see Rice (1989)). Finding further evidence to support this hypothesis will be left to future research.

If the *y*-prefix is not a postposition, then the *bi*-prefix cannot be the object of a postposition. The remainder of the paper will focus on the unexpected appearance and behavior of *bi*-object morphology, which we will argue is not postpositional agreement, contrary to Hale (1997).<sup>8</sup>

### 3. *Bi*-object agreement in unergative causatives

We begin by examining how basic 3rd person agreement works in Navajo.

#### 3.1. 3rd person agreement in Navajo

Normally, the 3rd person direct object is represented by  $\emptyset$  when the subject of the verb is other than 3rd person. When both subject and direct object are 3rd person, the 3rd person direct object must be represented by *yi*- or *bi*-. (See Young and Morgan 1987:64.) *Yi*- is the 3rd person object prefix required in normal SOV

---

<sup>8</sup>Young and Morgan’s (1987) analysis also supports the assertion that *bi*- is not the object of a postposition. They state (1987:65) that the prefix *bi*- which is used in transitivized intransitive verbs (i.e. unergative causatives) is the 3rd person direct object (position 4-conjunct), citing the following examples:

- |       |                                               |                                                                            |                            |
|-------|-----------------------------------------------|----------------------------------------------------------------------------|----------------------------|
| (i)   | <i>habiishyeed</i><br>cf. <i>haashyeed</i>    | ‘I’m running it up out (as a horse from a canyon)’<br>‘I’m running up out’ | (Young and Morgan 1987:65) |
| (ii)  | <i>habiishchxééh</i><br>cf. <i>haashchééh</i> | ‘I’m honking it (a car horn)’<br>‘I’m starting to cry’                     | (Young and Morgan 1987:65) |
| (iii) | <i>biisíʔ</i><br>cf. <i>yiizíʔ</i>            | ‘I stood him up’<br>‘I stood up’                                           | (Young and Morgan 1987:64) |

They add that “*Bi*-IV must not be confused with *bi*-0, the object of a postposition” (Young and Morgan 1987:65).

sentences when both subject and object are 3rd person. This is summarized in (13).

(13)

| If subject is | 3rd p. obj agreement is                      |
|---------------|----------------------------------------------|
| 1             | ∅                                            |
| 2             | ∅                                            |
| 3             | usually <i>yi-</i> ,<br>sometimes <i>bi-</i> |

The problem involves two separate issues:

- 1) The unergative causative sentences we have looked at so far have 1st person subjects. *Why is there any overt 3rd person object agreement at all?*
- 2) The usual object agreement when there are two 3rd persons is *yi-*. *Why is the object agreement *bi-* and not *yi-* in causativized unergatives?*

### 3.2. Question 1: 3rd person agreement with 1st/2nd subject

Many recent analyses argue that causatives undergo a type of argument sharing, whereby the internal argument of the causative predicate is semantically identified with the subject of the embedded predicate. In one such analysis, Alsina (1997) argues that causative complex predicates are formed by predicate composition in which an incomplete predicate is forced to combine with another predicate in order to complete its argument structure. This is illustrated by the causativization of the verb *laugh* in (14)-(17). The base predicate *laugh* has one (external) argument, which is a Proto-Agent. The External Argument Mapping Principle (Alsina 1997:207) requires that the external argument map on to the syntactic function of SUBJECT, as shown in (14).



The causative predicate, illustrated in (15), results from the morphological combination of the causative morpheme with a verb stem, with consequent composition of the predicate information (Alsina 1997:211). P\* followed by an underspecified a(rgument)-structure represents any predicator and its a-structure. This is an incomplete predicate that must compose with another predicate in order to be complete. The line connecting the argument of the causative predicate with an argument of the embedded predicate indicates that they are semantically identified, i.e. the same semantic participant (Alsina 1997:211).



**3.3. Question 2: *bi-* agreement marker when 3rd subject/3rd object**

We can now turn to our second question. The usual object agreement when there are two 3rd persons is *yi-*. Why is the object agreement *bi-* and not *yi-*?

To answer this question, we must digress a little and examine the instances where *bi-* object morphology is present. *Bi-* is the object prefix that shows up in the *yi-/bi-* alternation known as Subject/Object Inversion or the Inverse Construction. This construction is one of the most-discussed topics in the Athapaskan syntactic literature (see Hale 1973, Creamer 1974, Platero 1974, Perkins 1978, Sandoval 1984, Sandoval & Jelinek 1989, Willie 1989, 1991, Speas 1990, Thompson 1996, Uyechi 1996, Jelinek 1997, Horseherder 1998, Saxon and Rice 2001, and others.) Examples of the *yi-/bi-* alternation are given in (19)-(21).

- (19) a.    ʌɪɔ̃ɪɔ̃ɔ̃'    dzannez    **yi-**ztau  
           horse       mule       him-kicked  
           'The horse kicked the mule.'           (Hale 1973:300)
- b.    dzannez    ʌɪɔ̃ɪɔ̃ɔ̃'    **bi-**ztau  
           mule       horse       him-kicked  
           'The mule was kicked by the horse.'<sup>11</sup> (Hale 1973:300)  
           i.e. 'The horse kicked the mule' or  
           'The mule, the horse kicked him.'
- (20) a.    'ashkii    'at'ɛɛd    **yii**utsā/  
           boy       girl       3O-3SGS-saw  
           'The boy saw the girl.'           (Hale 1973:301)
- b.    'at'ɛɛd    'ashkii    **bii**utsā/  
           girl       boy       3O-3SGS-saw  
           'The boy saw the girl.'           (Hale 1973:301)
- (21) a.    ʌɛɛchãã'ɪ    mósɪ       **yish**xash  
           dog       cat       3O-3SGS-bit  
           'The dog bit the cat.'           (Hale 1973:301)

---

<sup>10</sup>The requirement that one of the arguments be an agent is redundant in the case of causativized unergatives. However, the relevance of this requirement is evident in the behavior of the causativized unaccusatives, where the two 3rd person arguments are both patient, and overt object marking does not occur. A very similar restriction holds in Tzotzil (a Mayan language) where the Agent Focus form (a type of Inverse) is only permitted in clauses with 3rd person agent and patient (Aissen 1999). The following section will show how the causative construction parallels the Inverse Construction in Navajo.

<sup>11</sup>It is important to note that although the *bi-* sentences are often given a passive gloss in English, this is an active sentence and not a passive one; an independent passive construction exists in Navajo.



*Object Marking and Agentivity in Navajo Causatives*

Each of these analyses of the Inverse Construction underscores the fact that *bi-* marks the subject as a patient. Without disputing this, we would like to propose that a slightly different perspective more accurately characterizes the function of *bi-*. Crucially, *bi-* refers to focus rather than topic. *Bi-* marks the object or “focus” position as agent.

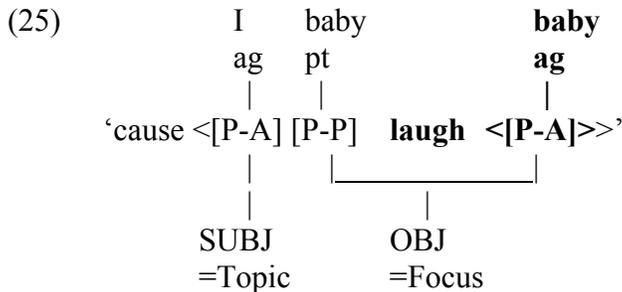
This is illustrated in (23). Adopting the topic/focus discourse terminology, “topic” refers to NP<sub>1</sub>, and “focus” refers to NP<sub>2</sub> in a sequence NP<sub>1</sub>-NP<sub>2</sub>-VERB.<sup>13</sup>

|      |            |                       |                       |            |                       |                       |
|------|------------|-----------------------|-----------------------|------------|-----------------------|-----------------------|
| (23) | <i>yi-</i> | AGENT                 | PATIENT               | <i>bi-</i> | AGENT                 | PATIENT               |
|      |            | T <small>OPIC</small> | F <small>OCUS</small> |            | T <small>OPIC</small> | F <small>OCUS</small> |

Canonically, agent maps to topic and patient maps to focus. These are the cases where *yi-* appears when there are two 3rd person arguments. However, in the Inverse Construction, a 3rd person *agent* maps to focus (and topic is a non-agent of the predicate).

How does this relate to our discussion of unergative causatives? Just as in the Inverse Construction, the *bi-* prefix indicates non-canonical mapping where *agent maps to focus* (and topic is a non-agent of the base predicate). This is illustrated in (24) and (25), repeated from above.

|      |       |                          |                                    |                |
|------|-------|--------------------------|------------------------------------|----------------|
| (24) | Topic | <b>Focus</b>             |                                    |                |
|      | (Shí) | 'awéé'                   | biyeeshdloh. (< bi-y-gh-sh-u-dloh) | (Hale 1997:53) |
|      | (I)   | <b>baby</b>              | 3-y-PROG:1s-u-d:laugh:PROG         |                |
|      |       | 'I make the baby laugh.' |                                    |                |



The *bi-* prefix indicates that *baby*, in *focus* position, is *agent* of the base predicate, rather than patient. As for the topic, *I*, *bi-* can only tell us that it is a non-agent of the base predicate *laugh*; this does not preclude it from being an agent of the causative predicate. It is not enough to say that *bi-* indicates a topical-

---

<sup>13</sup>We use the terms topic/focus rather than subject/object due to the conflicting uses of the terms subject/object in previous analyses. For example, while some analyses treat the *yi-/bi-* alternation as an SOV-OSV alternation, others do not.

ized patient; while this will explain the Inverse cases, it will not explain the causative cases nor the one-nominal cases.

Finally, this answers our second question: when there are two 3rd person arguments, the 3rd person object marker *bi-* is used rather than *yi-* to indicate that the focused NP is an agent of the base predicate.

#### 4. Conclusions

We have seen that two types of morphological causatives can be found in Navajo: those formed from unaccusative verbs and those from unergative verbs. The causativized unergative verbs exhibit several unique properties, including an additional “causative” prefix and the *bi-* prefix marking 3rd person object agreement, which is not the expected agreement.

Normally, overt 3rd person morphology is only required when there are two 3rd person arguments. By appealing to an analysis of complex predicates in which the internal argument of the causative predicate is semantically identified with the logical subject of the embedded predicate, we can explain why 3rd person object marking is necessary in causatives: there are two 3rd person arguments.

Secondly, the usual agreement found with two third person arguments is the *yi-* prefix. In causatives, however, the agreement prefix present is *bi-*. By examining the behavior of the *bi-* object agreement more closely, we can explain its unexpected appearance: the *bi-* prefix indicates a non-canonical mapping between argument structure and discourse structure in which agent maps to focus. A reinterpretation of the principles governing 3rd person object marking in Navajo thus makes it possible to unify two seemingly unrelated uses of the *bi-* prefix.

#### References

- Aissen, Judith. 1999. Agent focus and inverse in Tzotzil. *Language* 75: 451-485.
- Alsina, Alex. 1997. Causatives in Bantu and Romance. In Alex Alsina, Joan Bresnan and Peter Sells, eds., *Complex Predicates*, 203-246. Stanford: CSLI Publications.
- Alsina, Alex, Joan Bresnan and Peter Sells, eds. 1997. *Complex Predicates*. Stanford: CSLI Publications.
- Bittner, Maria and Kenneth Hale. 1996. The structural determination of case and agreement. *Linguistic Inquiry* 27: 1-68.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Creamer, Mary Helen. 1974. Ranking in Navajo nouns. *Dine Bizaad Nanil'ih/Navajo Language Review* 1: 29-38.
- Hale, Kenneth. 1973. A note on subject-object inversion. In B. Kachru, ed., *Issues in Linguistics. Papers in Honor of Henry and Renee Kahane*, 300-310. Urbana: University of Illinois Press.
- Hale, Kenneth and Paul Platero. 1996. Navajo reflections of a general theory of lexical argument structure. In Jelinek, Midge, Rice and Saxon, eds., *Atha-*

*Object Marking and Agentivity in Navajo Causatives*

- paskan Language Studies. Essays in Honor of Robert W. Young*, 1-14. Albuquerque: University of New Mexico Press.
- Hale, Kenneth. 1997. Remarks on the syntax of the Navajo verb. Ms. MIT and Navajo Language Academy Linguistics Workshop.
- Horseherder, Nicole. 1998. A binding-theoretic analysis of Navajo possessor *yi-*. M.A. thesis, University of British Columbia.
- Jelinek, Eloise. 1997. Topic and focus in Navajo inverse. In B. Agbayani and S-W. Tang, eds., *Proceedings of WCCFL XV*. Stanford: CSLI Publications.
- Kulikov, Leonid. 1993. The "2nd causative": A typological sketch. In B. Comrie and M. Polinsky, eds., *Causatives and Transitivity*, 121-154. Philadelphia: John Benjamins Publishing Company.
- McDonough, Joyce. 1990. *Topics in the Phonology and Morphology of Navajo Verbs*. Ph.D. dissertation, University of Massachusetts, Amherst.
- Oh, Sunyoung. 2000. Causatives and functional head-key in Korean. Ms., University of British Columbia.
- Perkins, Ellavina. 1978. The role of word order and scope in the interpretation of Navajo sentences. Ph.D. diss., University of Arizona.
- Platero, Paul. 1974. The Navajo relative clause. *International Journal of American Linguistics* 40: 202-246.
- Rice, Keren. 1989. *A Grammar of Slave*. Berlin: Mouton de Gruyter.
- Rice, Keren. 1991. Intransitives in Slave (Northern Athapaskan): Arguments for unaccusatives. *International Journal of American Linguistics* 57: 51-69.
- Rice, Keren. 2000a. *Morpheme Order and Semantic Scope*. Cambridge: Cambridge University Press.
- Rice, Keren. 2000b. Monadic verbs and argument structure in Ahtna, Slave and Navajo. In T. Fernald and P. Platero, eds., *Athabaskan Language and Linguistics*, 167-199. Oxford: Oxford University Press.
- Sandoval, Merton. 1984. The syntactic function of the *yi/bi-* alternation in Jicarilla Apache. *Coyote Papers*, vol. V. Tucson: University of Arizona.
- Sandoval, Merton & Eloise Jelinek. 1989. The *bi-* construction and pronominal arguments in Apachean. In Cook and Rice, eds., *Athapaskan Linguistics: Current Perspectives on a Language Family*, 335-378. Berlin: Mouton de Gruyter.
- Saxon, Leslie and Keren Rice. 2001. A working paper on two OSV constructions in Navajo: Some data to start with. Handout, University of Victoria Linguistics Workshop.
- Speas, Margaret. 1990. *Phrase Structure in Natural Language*. Dordrecht: Kluwer Academic Publishers.
- Thompson, Chad. 1996. The history and function of the *yi-/bi-* alternation in Athapaskan. In Jelinek, Midgette, Rice and Saxon, eds., *Athapaskan Language Studies. Essays in Honor of Robert W. Young*, 81-100. Albuquerque: University of New Mexico Press.

Suzanne Gessner

- Uyechi, Linda. 1996. The Navajo 3rd person alternation and the pronoun incorporation analysis. In Jelinek, Midgette, Rice and Saxon, eds., *Athapaskan Language Studies. Essays in Honor of Robert W. Young*, 123-136. Albuquerque: University of New Mexico Press.
- Willie, Mary Ann. 1989. Why there is nothing missing in Navajo relative clauses. In Cook and Rice, eds., *Athapaskan Linguistics: Current Perspectives on a Language Family*, 407-438. Berlin: Mouton de Gruyter.
- Willie, Mary Ann. 1991. Pronouns and obviation in Navajo. PhD dissertation, University of Arizona.
- Young, Robert and William Morgan. 1987. *The Navajo Language*. Albuquerque: University of New Mexico Press.
- Young, Robert and William Morgan (with Sally Midgette). 1992. *Analytical Lexicon of Navajo*. Albuquerque: University of New Mexico Press.

Department of Linguistics  
Buchanan E270, 1866 Main Mall  
University of British Columbia  
Vancouver, B.C. V6K 1E7

[gessner@interchange.ubc.ca](mailto:gessner@interchange.ubc.ca)

# An Account of Compensatory Lengthening without Moras

BEN HERMANS  
*Tilburg University*

## 0. Introduction

Hayes' (1989) theory of Compensatory Lengthening (CL) makes two closely related predictions. 1) CL is the result of mora preservation; and 2) the loss of a segment from an onset position cannot lead to CL. Hayes' theory is embedded in a derivational conception of phonology. It might be worthwhile, then, to investigate if, and to what extent, these two predictions follow from a non-derivational theory.

It turns out that we are forced to give up the two predictions if we follow the basic premises of Optimality Theory (OT). Interestingly, as far as the first prediction is concerned this is not a disappointing result. In fact, Slovak has a type of CL that has nothing to do with mora preservation. In this respect, then, OT's suspicion towards the rule based theory of CL seems justified. However, to give up the second prediction is not desirable, because convincing cases where the loss of a segment from an onset position triggers CL have never been found. It is the main goal of this article to develop a theory of CL in OT that allows us to maintain the second prediction.

I propose that CL has nothing to do with mora preservation. Basically, it is segment preservation but with one important addition; the output segment and the input segment must be very close in terms of sonority.

In section 1, I argue that CL has nothing to do with mora preservation, a conclusion that is very welcome from the perspective of OT. In section 2, I propose that CL really is the result of segmental faithfulness, a relation that is only possible if the corresponding segments are identical in terms of sonority in a specific sense. It is this identity requirement which explains why the loss of an onset segment is not compensated for.

## 1. CL: mora preservation or segment preservation?

In this section I proceed as follows. First, I briefly sketch Hayes' theory of CL. The central generalization of this theory is that CL is the result of mora preservation. Then I will argue that from the perspective of OT there are reasons to be

suspicious about this generalization, simply because it cannot be expressed. Then I show that this suspicion is justified because there are indeed instances of CL where mora preservation is not relevant at all.

### 1.1. CL is mora preservation; the classical view

According to Hayes, CL is what you get when a segment is deleted but its mora is maintained. The preservation of the mora triggers the spreading of an adjacent segment because the mora that is left behind is filled. This is shown in (1) in a schematic form. (1a) is an instance of what Hayes calls ‘Classical CL,’ where a consonant in coda position is dropped and replaced by a lengthened vowel (if the stray mora is filled by the adjacent vowel) or a lengthened consonant (if the stray mora is filled by the adjacent consonant). (1a) only represents the first option. (1b) exemplifies what Hayes calls ‘Double Flop.’ In this type of CL, a segment in onset position is deleted. Then a neighboring segment occupies the onset position. In doing so it flees from its original position, leaving its mora stray. This mora is then filled by the vowel or the consonant. (1b) only represents the first option. In (1) syllable structure is indicated by dots and by spacing.<sup>1</sup>

(1) a. *Classical CL*

|                 |              |                   |           |
|-----------------|--------------|-------------------|-----------|
| syllabification | segment loss | resyllabification | spreading |
| mm m            | m m m        |                   | m m m     |
| .c v c .c v     | .c v .c v    |                   | .c v .c v |

b. *Double Flop*

|                 |              |                   |           |
|-----------------|--------------|-------------------|-----------|
| syllabification | segment loss | resyllabification | spreading |
| mm m            | m m m        | m m m             | m m m     |
| .c v c .c v     | .c v c .v    | .c v .c v         | .c v .c v |

Examples of Classical CL are cases like Latin *fide:lia* ‘pot’ and *ca:nus* ‘old’, deriving from older *\*fideslia* and *\*kasnus*. An example of Classical CL illustrating lengthening of the adjacent *consonant* is the underlying Pali form */kar+ta/* ‘make’ realized as *katta*. Examples illustrating Double Flop (1b), where the loss of the consonant triggers vowel lengthening, are Doric *kha:nos* ‘goose’ (gen. sg.) and *este:la* ‘I sent’ developed from earlier *\*khansos* and *\*estelsa*. An example of Double Flop, where the loss of the consonant is compensated for by *consonantal* lengthening is the Pali form */lag+na/*, realized as *lagga* (the Greek examples are taken from De Chene and Anderson (1979); cf. Wetzels (1986) for an analysis of CL in Greek dialects in terms of Double Flop; the Pali examples are from Zec (1995)).

<sup>1</sup> The schematic examples in (1) are just two instances of CL. For a full typology I refer to Hayes’ article.

The two patterns in (1) illustrate the essence of Hayes' theory. In these configurations, segment deletion affects only the segmental layer not the mora level. Thus, after the deletion of the consonant, its mora is left behind. It is subsequently filled by spreading, creating a long vowel or a long consonant. In sum, after deletion the number of moras is the same as before deletion. That is why CL is a consequence of mora preservation, according to Hayes. Let us now look at this hypothesis from the point of view of OT.

### **1.2. Is CL really mora preservation? OT's suspicion**

The claim that CL is the consequence of mora preservation obviously implies that moras must be present *before* the relevant segment is removed. Hayes' theory of CL is embedded in a *derivational* conception of phonology, so it is quite easy to guarantee that moras are inserted before the segmental tier is affected. The theory just declares that syllabification is an 'anywhere' rule, which applies whenever it gets a chance. This means that mora construction applies to a string right after it leaves the lexicon. Consequently, moras will always be present *before* any phonological rule has a chance to apply.

In OT there are no rules (or constraints) that apply before or after some other rule or constraint, because in OT there are no derivations. As a result of the non-derivational nature of OT it is difficult to *ensure* that moras are present before segments are deleted. Consider the Pali form /*nud+na*/ 'remove' (past. part.), realized as *nunna*. The infinitive is /*nud+ati*/ realized as *nudati*. Clearly, the final consonant of the root /*nud*/ cannot be moraic underlyingly, because then it would appear as a geminate in the infinitive, giving \**nuddati*. But if the consonant is underlyingly non-moraic, then it becomes hard to understand what the source is of the length in the past part. Why is the root consonant not simply deleted? Why does its deletion trigger lengthening of the following consonant? What we need is an *intermediate* level where moras are inserted but coda consonants are unaffected. But given its nature such a level cannot exist in OT.

From the point of view of OT, then, there is some reason to be skeptical about Hayes' hypothesis that CL is the result of mora preservation, simply because in OT this generalization cannot easily be expressed. At first this might seem something to worry about, so presumably one's first inclination would be to restructure OT, for instance by allowing intermediate levels in some form. On closer view, however, it turns out not to be a bad result, because there are cases of CL where the mora does not play a role at all. This will be shown in the next subsection.

### **1.3. CL is not mora preservation**

In his detailed analysis of Slovak, Rubach (1993) shows that the loss of a so called 'yer' triggers lengthening of a preceding vowel. Consider the following examples (a yer is indicated by a capital letter):

|     |                            |                      |                         |
|-----|----------------------------|----------------------|-------------------------|
| (2) | <i>basic form, nom. sg</i> | <i>dim. nom. sg.</i> | <i>dim. gen.</i>        |
|     | hlas ‘voice’               | hlas+Ok+O [hla:sok]  | hlas+Ok+a [hla:ska] sg. |
|     | sud ‘cup’                  | sud+Ok+O [su:dok]    | sud+Ok+a [su:dka] sg.   |
|     | hlav+a ‘head’              | hlav+Ok+a [hla:vka]  | hlav+Ok+O [hla:vok] pl. |

In the dim. nom. sg. of the masc. paradigm (the first two examples in (2)) and the gen. plur. of the fem. paradigm there are two yers in the underlying form. The first one is removed and replaced by a lengthened vowel in the preceding syllable. The second yer is realized in preconsonantal position. In the dim. of the gen. masc. sg. and the nom. sg. of the fem. paradigm just one yer is present in the underlying representation. This yer is removed and replaced by a lengthened vowel in the preceding syllable. In this article we are not concerned with the rules regulating the distribution of yers, of course. What is important to us here is the fact that if a yer is deleted its loss is compensated for by lengthening of the preceding vowel. This indeed strongly suggests that this form of lengthening is a case of CL.

Now the point is that according to the dominant view a yer is a mora-less vowel in the underlying representation. This position has been taken by many authors (in particular Rubach (1993) on Slovak). If this is true, then it is clear that Hayes’ hypothesis can no longer be maintained. To see this more clearly compare the underlying form of *little head* with its surface realization.

|     |                        |                               |
|-----|------------------------|-------------------------------|
| (3) | <i>underlying form</i> | <i>surface representation</i> |
|     | m m                    | m m m                         |
|     | hl a v o k a           | hl a v k a                    |

The underlying form has just two moras, whereas the surface representation has three. This shows that in Slovak, CL is not a consequence of mora preservation.

From the point of view of OT this is not really a surprise, because Hayes’ hypothesis cannot easily be expressed in OT, as shown before. Having shown that Hayes’ theory is questionable both on theory internal grounds (OT) and on descriptive grounds (Slovak), I will try to suggest an alternative approach to CL.

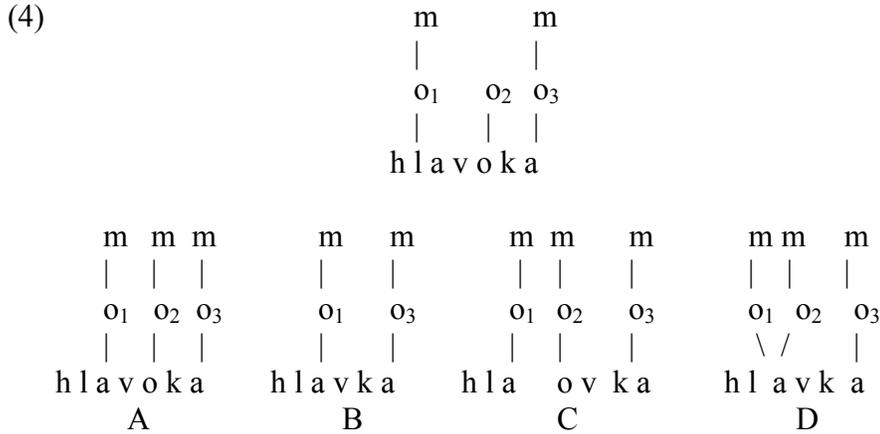
## 2. An alternative approach to CL

In this section, I first propose that CL is an instance of segmental faithfulness. Then I will make an attempt to capture Hayes’ second major insight; the loss from onset position cannot directly lead to CL (cf. (1)).

### 2.1. CL as segmental faithfulness

Consider again the underlying representation of *little head* but now in a slightly more formal version, including not only moras but also vocalic root nodes. This underlying form is located in the upper row of (4). The lower row contains several

surface candidates corresponding to this underlying form. The one at the right end (D) is the optimal candidate.



Of course there is not sufficient room to give a detailed analysis of the Slovak type of CL. But the representations already suffice to make the following point. It is possible to account for CL with the apparatus of segmental faithfulness. To make this explicit I have indicated satisfaction of MAX-seg<sup>2</sup> with subscripts. Candidate B is rejected because its second segment, the yer, does not satisfy MAX-Seg. Furthermore, it is clear that an underlying yer cannot be realized in the head position of a syllable, due to the constraint HEAD-DEP-m.<sup>3</sup> That is why the first candidate is non-optimal. This leaves us with the remaining two candidates. In candidate C, the underlying yer is realized in the non-head position, so that is good. However, by doing so a diphthong is realized. On the assumption that a violation of NODIPH<sup>4</sup> is also bad in Slovak the third candidate is also rejected. This leaves us with the optimal candidate. In this form all the important constraints of Slovak are satisfied, in particular MAX-seg.

Notice that in the optimal candidate IDENT(F) is violated.<sup>5</sup> Here we arrive at the essential point. It is possible to analyze CL with the apparatus of segmental faithfulness. What we need is non-violation of MAX-seg, and large scale violation of IDENT(F), due to some higher constraint conflicting with it (in Slovak, NO-

<sup>2</sup> This constraint requires that an underlying segment (root node) correspond to a surface segment.  
<sup>3</sup> This constraint penalizes the insertion of a mora in the head position of a syllable.  
<sup>4</sup> This constraint penalizes a diphthong. It does not necessarily imply that Slovak does not allow diphthongs; it depends on the position of NODIPH in the hierarchy. In fact, lengthened *e* and *o* are realized as *ie* and *uo*. The point is that *certain* diphthongs (like *ao*) cannot be produced by lengthening. Naturally, a full account has to take all these aspects of lengthening into consideration. Within the limited space of this article, however, this is impossible.  
<sup>5</sup> This constraint penalizes qualitative differences between two corresponding segments.

DIPH).<sup>6</sup> In sum, in OT there is an easy way to account for CL by means of the following ranked constraints:

- (5) *CL as segmental faithfulness*  
 MAX-seg, C » IDENT(F)

In order to get CL, MAX-seg must be ranked above IDENT(F). In addition, there must be some other constraint conflicting with IDENT(F) and dominating it.

So far, I have shown that OT has reasons to be skeptical about Hayes' claim that CL is a consequence of mora preservation. Interestingly, it turns out that instances of CL contradicting Hayes' claim do exist. Finally, I have shown that, in principle, it is perfectly possible to account for CL in terms of segmental faithfulness. The constraint hierarchy in (5) clearly demonstrates that you can get CL effects without moras.

This might all be very well, but one problem remains. We still have to answer the question of whether Hayes' second generalization can be captured in a theory of CL that relies on segmental faithfulness. This will be the subject of the next subsection.

## 2.2. Restructuring the root node

Hayes' second important claim is that the loss of a segment from an onset cannot be compensated for.<sup>7</sup> Here are the two crucial cases in a schematic form:

- (6) a. *initial onset*
- |                 |              |
|-----------------|--------------|
| syllabification | segment loss |
| m               | m            |
| c c v           | c v          |
- b. *intervocalic onset*
- |          |        |
|----------|--------|
| m m      | m m    |
| c v . cv | c v .v |

(6a) shows in a schematic form that onset simplification cannot lead to a long vowel, because after the loss of the segment occupying the onset, no trace is left behind. Similarly, intervocalic weakening with ultimate deletion of the entire segment, the case depicted in (6b), never leads to lengthening of the adjacent vowel. These generalizations are robust, and they are explained in Hayes' theory

<sup>6</sup> I should note that it is necessary to assume that length is represented with two root nodes, as proposed in Selkirk (1991).

<sup>7</sup> Recall from (1b) that Double Flop is not a counterexample. Although there is an onset segment being lost and although there is also another segment being lengthened, it really is the shift of the original coda segment to the onset that causes lengthening.

in a straightforward way. In the relevant positions, segments do not have a mora. If they are deleted from such a position, they disappear once and for all.

It is immediately clear that an account of CL in terms of segmental faithfulness cannot express this generalization. This is shown in (7), where the subscripts represent the relation holding between underlying segments and their correspondents in the output (as far as it is relevant to the discussion here).

- (7) a. *initial onset*
- |                                 |                                 |  |
|---------------------------------|---------------------------------|--|
| underlying                      | segment loss accompanied by     |  |
| form                            | vowel lengthening               |  |
| m                               | m m                             |  |
| c c <sub>1</sub> v <sub>2</sub> | c v <sub>1</sub> v <sub>2</sub> |  |
- 
- b. *intervocalic onset*
- |                                                 |                                                  |    |                                                |
|-------------------------------------------------|--------------------------------------------------|----|------------------------------------------------|
| underlying                                      | segment loss accompanied by                      |    |                                                |
| form                                            | vowel lengthening                                |    |                                                |
|                                                 | to the left                                      |    | to the right                                   |
| m m                                             | m m m                                            |    | m m m                                          |
| c v <sub>1</sub> .c <sub>2</sub> v <sub>3</sub> | c v <sub>1</sub> v <sub>2</sub> . v <sub>3</sub> | or | cv <sub>1</sub> .v <sub>2</sub> v <sub>3</sub> |

These configurations show that the ranking of (5) can easily lead to a situation in which the loss of a segment from an onset position leads to lengthening of an adjacent vowel. This surely is a very undesirable result indeed.

While this conclusion is correct in itself, I would like to suggest that this does not necessarily mean that we have to take refuge in the mora again. In fact, there are indications that this would be a false move. In a very important paper, Rialland (1993) argues quite convincingly that the mora theory of CL, or any other theory for that matter, does not explain the following characteristic of CL phenomena:

seuls les segments les plus sonorants .... peuvent donner lieu à un allongement compensatoire et seule leur chute peut laisser une unité chronématique de sonorance suffisante pour être remplie par une voyelle (91-2).<sup>8</sup>

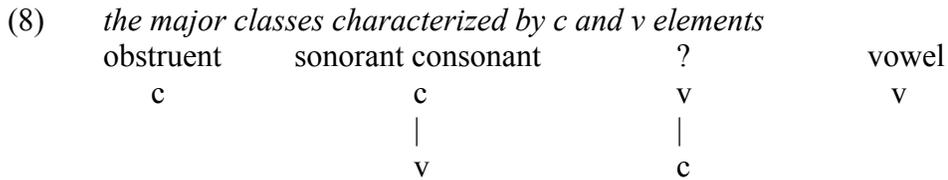
The importance of this remark can hardly be underestimated, for it explicitly claims that it is not really moras that matter but rather identity at the level of sonority. If we restate it in OT terminology, we could say that somehow the input segment (that is going to be deleted) should be very close in terms of sonority to the output segment (by which it is going to be replaced). When we look at Rialland's remark from this angle it is clearly reminiscent of segmental faithfulness,

---

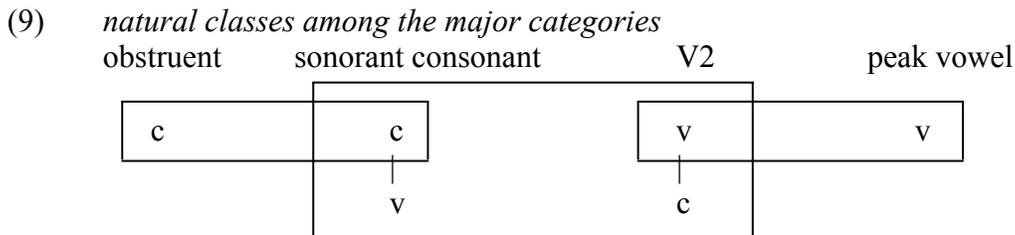
<sup>8</sup> "only the most sonorant segments....can give room to compensatory lengthening and only their fall can leave behind a timing unit that is sufficiently sonorant to be filled by a vowel" (my translation).

which is precisely the track we are pursuing here. How then can we implement Rialland's insight at a more formal level, i.e. in terms of faithfulness constraints?

First of all, following the lead of Anderson and Ewen (1987), I propose that the major classes are distinguished by c- and v-elements, c indicating a relatively high degree of periodic energy and v a relatively low degree of periodic energy. Both elements constitute the root node, but if they are combined in a single segment one is dependent on the other. In the spirit of this proposal one might expect the following major classes:



It is clear that an obstruent is just a c at the root node. It is also clear that a vowel is just a v. It is also reasonable to assume that a sonorant consonant is a c-element with a dependent v. But what is the third logically possible element? I propose that this feature combination is the representation of the second half of a long vowel. This representation is not unreasonable, since, as is well known, the second half of a long vowel tends to be reduced in many languages. If the categories of (8) are acceptable then we can take the next step. Look at the natural classes that can be defined by the features in (8):



Sonorant consonants can pattern with obstruents, because both categories have a c-head. They can also pattern with the second half of a long vowel, because these two classes are identical at the level of the entire root node; the only difference between these two classes is the dependency relation between the two root node features. Finally, the third (and from our perspective least interesting) natural class consists of the two types of vowels.

The next move consists of a reformulation of MAX-seg.

- (10) **MAX-seg**  
 A root node feature in the input should correspond to an identical feature in the output.

This formulation is based on the idea that correspondence plays a role at the feature level not the segmental level. According to the formulation in (10), it is the case that for every *c* in the input there must be a *c* in the output, and likewise, for every *v* in the input there must be a *v* in the output. Notice, however, that nothing is said about the *dependency relation* between elements. Consequently, as far as MAX-seg is concerned, the dependency relation in the domain of a root node can freely be changed. It just requires that a root node's substantial content remain the same. Let us now see what the consequences are of this new formulation of MAX-seg.

Surely, the most important consequence is that a segment of the input can only correspond to a segment of the output if both segments are very close in terms of sonority. This is a desirable consequence, because this is the most characteristic property of CL, according to Rialland (1993). Given the theory of the root node proposed in (8), this means that an underlying sonorant consonant can only correspond to a vowel if that vowel is the *second half of a long vowel*. It can never correspond to a vowel in peak position, because a sonorant consonant and a peak vowel are not members of the same major category. It is this property of MAX-seg that gives us the opportunity to maintain Hayes' second hypothesis.

To see this more clearly let us go back to (7), repeated immediately below, where all possible cases of CL triggered by onset loss are represented in a schematic form.

- (11) a. *initial onset*
- |                                 |                                 |  |
|---------------------------------|---------------------------------|--|
| underlying                      | segment loss accompanied by     |  |
| form                            | vowel lengthening               |  |
| m                               | m m                             |  |
| c c <sub>1</sub> v <sub>2</sub> | c v <sub>1</sub> v <sub>2</sub> |  |
- b. *intervocalic onset*
- |                                                 |                                                  |    |                                                |
|-------------------------------------------------|--------------------------------------------------|----|------------------------------------------------|
| underlying                                      | segment loss accompanied by                      |    |                                                |
| form                                            | vowel lengthening                                |    |                                                |
|                                                 | to the left                                      |    | to the right                                   |
| m m                                             | m m m                                            |    | m m m                                          |
| c v <sub>1</sub> .c <sub>2</sub> v <sub>3</sub> | c v <sub>1</sub> v <sub>2</sub> . v <sub>3</sub> | OR | cv <sub>1</sub> .v <sub>2</sub> v <sub>3</sub> |

The first case (11a) might be possible in a traditional account of segmental faithfulness, but it is no longer possible in the new version proposed here. In (11a), the onset consonant corresponds to a peak vowel. According to the new proposal developed here this is not possible, because these two segment types are not sufficiently close in terms of sonority, or more formally: a vowel in peak position cannot stand in a correspondence relation to a (sonorant) consonant.

The same can be said about the second case depicted in (11b). The loss of an onset consonant cannot yield a long vowel to its right because MAX-seg cannot establish a relation between a (sonorant) consonant and a peak vowel.

We are thus left with the first case given in (11b). Here a (sonorant) consonant corresponds to the second half of a long vowel. According to the new version of MAX-seg this is a possible relation. Hence a CL process of this type should be possible. Although it might be true that the system of segmental faithfulness developed here might be able to describe a process in which the loss of an intervocalic onset triggers lengthening of a preceding vowel, there are other reasons why a process of that kind is highly unlikely. Normally, the loss of a consonant from an intervocalic position is the consequence of weakening. Obviously then we need a constraint accounting for the fact that intervocalic consonants undergo weakening (and ultimately deletion). Suppose we formulate this constraint as (12), where *c* and *v* are elements in the specific sense of this article.

(12) **INTERVOCALIC WEAKENING**

\**v c v*

According to this constraint, a segment carrying *c* is not allowed in between vowels. Notice now that it does not make sense to move the *c* to the second position of a long vowel. Such a repair strategy is useless, because in this position the *c* is still intervocalic. Consequently, the loss of intervocalic consonants, normally triggered by INTERVOCALIC WEAKENING, cannot be compensated for by lengthening of the preceding vowel. As a repair strategy it makes no sense.

I have argued that an account of CL in terms of segmental faithfulness can also make the prediction that the loss of a segment from onset position cannot be compensated for by vowel lengthening. It is necessary, however, to refine MAX-seg. It is also necessary to change our view of the root node. As a result we are led to the prediction that the loss of a segment can only lead to vowel lengthening if the input and output segments are sufficiently close in terms of sonority. This, in fact, is the property of CL that any theory should account for, according to Rialland.<sup>9</sup>

---

<sup>9</sup> The attentive reader might object that the new formulation of MAX-seg cannot account for those cases where an underlying sonorant corresponds to one half of a geminate obstruent. This type of CL exists, as is shown by the Pali form */kar+tun/* ‘to make’ (inf.) realized as *kattun*. The new formulation of MAX-seg cannot account for the correspondence relation between the input sonorant (*c+v*) and the first half of the geminate obstruent in the output (just *c*). One way to include total assimilation in the class of possible CL phenomena is to postulate a head version of MAX-seg. This constraint would essentially say the same thing as the constraint in (10), but it would restrict its application to the HEAD-element of the root node. It would thus read as follows: A HEAD-root node feature in the input should correspond to an identical feature in the output.

### **3. Conclusion**

In this paper, I have argued that the classical theory of CL proposed in Hayes (1989) is suspect from the point of view of OT, because it is derivational in nature. What is particularly disturbing is the role played by the mora. Interestingly, it turns out that Slovak has a case of CL in which the mora does not play a role at all. In this language, CL is clearly a case of segmental faithfulness; it has nothing to do with mora preservation. I have also shown that in an account of CL phrased in terms of segmental faithfulness it is still possible to maintain Hayes' second prediction that loss from an onset position cannot directly lead to vowel lengthening. This, however, is only possible if we change the formulation of MAX-seg and the structure of the root node. We are thus led to the following characterization of CL: CL is segmental faithfulness, but the corresponding segments should be sufficiently close in terms of sonority.

### **References**

- Anderson, J. M. and C. J. Ewen. 1987. *Principles of Dependency Phonology*. Cambridge: Cambridge University Press.
- De Chene, B. and S. R. Anderson. 1979. Compensatory Lengthening. *Language* 55: 505-35.
- Hayes, B. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20: 253-306.
- Rialland, A. 1993. L'allongement compensatoire: Nature et modèles. In B. Laks and A. Rialland, eds., *Architecture des représentations phonologiques*, 59-92. CNRS Editions.
- Rubach, J. 1993. *The Lexical Phonology of Slovak*. Oxford: Oxford University Press.
- Wetzels, L. 1986. Phonological timing in Ancient Greek. In L. Wetzels and E. Sezer, eds., *Studies in Compensatory Lengthening*, 297-344. Dordrecht: Foris.
- Selkirk, L. 1991. A two-root theory of length. Ms., University of Massachusetts, Amherst.
- Zec, D. 1995. The role of moraic structure in the distribution of segments within syllables. In J. Durand & F. Katamba, eds., *Frontiers of Phonology: Atoms, Structures, Derivations*, 149-180. London: Longman.

Tilburg University  
PO Box 90153  
5000 LE Tilburg  
The Netherlands

b.j.h.hermans@kub.nl



# Interactions between Constructions and Constraints in VP Ellipsis and VP Fronting\*

JONG-BOK KIM  
*Kyung Hee University*

## 0. Introduction

English VP Ellipsis (VPE) has been traditionally regarded as a process that maps a ‘syntactically complete’ sentence like (1a) onto a ‘syntactically incomplete’ one, like (1b), that is semantically equivalent to it (cf. Sag 1976).

- (1) a. John could leave before I could leave.  
b. John could leave before I could.

As pointed out in many places, one of the striking properties of VPE is that unlike phenomena such as gapping and comparative deletion, VPE does not obey sentence grammar rules (see Chao 1987, Johnson 2001, Lobeck 1995, Williams 1977, among others). Following are some of the main properties of VPE:

- VPE can appear across utterance boundaries:
- (2) A: Tom won’t leave Seoul soon.  
B: I don’t think Mary will \_\_ either.
- VPE can occur in either a subordinate or coordinate clause separate from the clause containing its antecedent:
- (3) a. Mary met Bill at Stanford although Sue didn’t \_\_ .  
b. Tom thinks that Mary met Bill at Stanford, but Sarah knows that Sue didn’t \_\_ .
- VPE obeys the Backwards Anaphora Constraint. As in (4), the ellipsis can precede, but not command, its antecedent (cf. Langacker 1966):

---

\* My deep thanks go to Emily Bender, Sae-Youn Cho, Chan Chung, Charles Fillmore, Andreas Kathol, Ivan Sag, Peter Sells, and the audiences of the 27th BLS for their valuable comments and criticisms. All errors and misinterpretations are, of course, mine. I also wish to acknowledge the financial support of the Korea Research Foundation (KRF-2000-041-A00255).

- (4) a. \*Sue didn't [e] but John ate meat.  
b. Because Sue didn't [e], John ate meat.
- VPE operates only on phrasal categories, as in (5).
- (5) a. \*Mary will meet Bill at Stanford because she didn't \_\_ John.  
b. Mary will meet Bill at Stanford because she didn't \_\_ at Harvard.
- VPE violates island constraints (cf. Napoli 1985):
- (6) a. John didn't hit a home run, but I know a woman who did \_\_ . (CNPC)  
b. That Betsy won the batting crown is not surprising, but that Peter didn't know she did \_\_ is indeed surprising. (SSC)  
c. Peter never hit a home run, but Betsy did and she was very happy about it. (CSC)
- VPE can induce ambiguity, implying that in determining the antecedent of an elided VP we must take into account its context.
- (7) Although Mary could have \_\_ , John decided to open the door. (*could have decided/could have opened*).
- VPE allows split antecedents, as illustrated in (8).
- (8) John was going to write a letter and Sue was going to send flowers, but one of them didn't \_\_ . (*write a letter and send flowers*).

Such properties have led Williams (1977) and others to treat VPE within Discourse Grammar differently from Sentence Grammar. Among several research issues in the analysis of VPE, this paper deals with the syntactic environment of VPE licensing within the framework of Head-Driven Phrase Structure Grammar (HPSG). In particular, it argues that language constraints on argument realization for auxiliary verbs and relevant constructions can provide a clean analysis of puzzling VPE and VP fronting phenomena.

## 1. VP Ellipsis

### 1.1. VPE in General Cases

The standard generalization of VPE is that it is possible only after an auxiliary verb, as shown in the contrast between (9) and (10).

- (9) a. Kim can dance, and Sandy can \_\_ , too.  
b. Kim has danced, and Sandy has \_\_ , too.  
c. Kim was dancing, and Sandy was \_\_ , too.

### VP Ellipsis and VP Fronting

- (10) a. \*Kim considered joining the navy, but I never considered \_\_ .  
b. \*Kim got arrested by the CIA, and Sandy got \_\_ , also.  
c. \*Kim wanted to go and Sandy wanted \_\_ , too.

The first issue in the analysis of VPE is the status of the elided VP. Following Lobeck (1995), López (2000), and Hardt (1999), among others, I take the elided VP phrase to be a *pro* element. The properties of the VPE we have seen in the beginning can be also found in pronouns. First of all, pronouns are phrases and can appear across utterance boundaries, as in (11). In addition, they can occur in coordinate or subordinate clauses, as in (12), are subject to the Backwards Anaphora Constraint, as in (13), can violate island constraints, as in (14), and can even have split antecedents, as in (15).

- (11) A: Does John eat fish?  
B: Yes, but **he** hates it.
- (12) John eats fish because/and **he** hates meat.
- (13) a. Because **he** doesn't like meat, John ate fish.  
b. \***He** doesn't like meat because John hates killing animals.
- (14) Bill really likes his new car. I think that the fact that **it** is an antique was a big selling point.
- (15) John arrived and later Susan arrived. **They** left together.

In accounting for the *pro*-drop phenomenon exemplified by Korean sentences like (16), we do not need to posit a phonologically empty pronoun if a level of argument structure is available (cf. Bresnan 1982, Bender 2000):

- (16) John-i *pro* *poassta*.  
John-NOM saw  
'John saw (him).'

We can simply encode the pronominal properties in the argument structure. For example, as represented within the feature structure of HPSG, illustrated in (17), the transitive verb *poassta* 'saw' takes a *pro* object NP as its argument, but the *pro* NP is not instantiated as the syntactic complement of the verb (as marked by the shaded area):

(17) *poassta* ‘see’

|             |                                                                                                  |
|-------------|--------------------------------------------------------------------------------------------------|
| <i>word</i> |                                                                                                  |
| HEAD        | $\left[ \begin{array}{l} \textit{verb} \\ \textit{VFORM} \quad \textit{fin} \end{array} \right]$ |
| SUBJ        | $\langle \boxed{1} \rangle$                                                                      |
| COMPS       | $\langle \quad \rangle$                                                                          |
| ARG-ST      | $\langle \boxed{1}\text{NP}, \text{NP}[\textit{pro}] \rangle$                                    |

Adopting this treatment of *pro* phenomena as a mismatch between the argument-structure and the syntactic valence features (cf. Manning and Sag 1999), we could interpret English VPE as a language-specific constraint of the argument realization constraint on *auxiliary verbs*, as represented in (18):

(18) Argument Realization Constraint on *aux-verbs*:

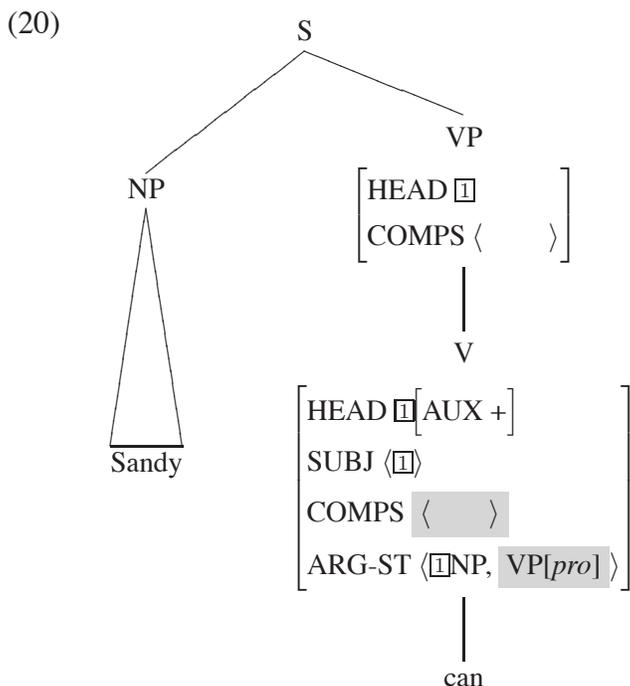
|                 |   |                                                                                                                                                                                                |
|-----------------|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>aux-verb</i> | → | $\left[ \begin{array}{ll} \text{SUBJ} & \boxed{A} \\ \text{COMPS} & \boxed{B} \\ \text{ARG-ST} & \boxed{A} \oplus \boxed{B} \oplus \textit{list}(\text{XP}[\textit{pro}]) \end{array} \right]$ |
|-----------------|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

What the constraint in (18) tells us is that when the final phrasal element in the argument-structure list of an auxiliary verb is a *pro*, this *pro* phrase need not be realized in the COMPS list, relevant to syntax. For example, the lexeme of the auxiliary verb *can* in (19a) takes a VP[*bse*] as its complement. When this VP is realized as a *pro* element, it need not appear in its COMPS list, as illustrated in (19b):

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                        |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|--|------|-----|------|-----------------------------|-------|-----------------------------|--------|------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|--|------|-----|------|-----------------------------|-------|-------------------------|--------|---------------------------------------------------------------|
| <p>(19) a. lexeme <i>can</i>:</p> <table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding-right: 10px;"><i>lexeme</i></td> <td></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 10px;">PHON</td> <td>can</td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 10px;">SUBJ</td> <td><math>\langle \boxed{1} \rangle</math></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 10px;">COMPS</td> <td><math>\langle \boxed{2} \rangle</math></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 10px;">ARG-ST</td> <td><math>\langle \boxed{1}\text{NP}, \boxed{2}\text{VP}[\textit{bse}] \rangle</math></td> </tr> </table> | <i>lexeme</i>                                                          |  | PHON | can | SUBJ | $\langle \boxed{1} \rangle$ | COMPS | $\langle \boxed{2} \rangle$ | ARG-ST | $\langle \boxed{1}\text{NP}, \boxed{2}\text{VP}[\textit{bse}] \rangle$ | <p>b. word <i>can</i> in VPE</p> <table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding-right: 10px;"><i>word</i></td> <td></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 10px;">PHON</td> <td>can</td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 10px;">SUBJ</td> <td><math>\langle \boxed{1} \rangle</math></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 10px;">COMPS</td> <td><math>\langle \quad \rangle</math></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 10px;">ARG-ST</td> <td><math>\langle \boxed{1}\text{NP}, \text{VP}[\textit{pro}] \rangle</math></td> </tr> </table> | <i>word</i> |  | PHON | can | SUBJ | $\langle \boxed{1} \rangle$ | COMPS | $\langle \quad \rangle$ | ARG-ST | $\langle \boxed{1}\text{NP}, \text{VP}[\textit{pro}] \rangle$ |
| <i>lexeme</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                        |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| PHON                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | can                                                                    |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| SUBJ                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | $\langle \boxed{1} \rangle$                                            |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| COMPS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | $\langle \boxed{2} \rangle$                                            |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| ARG-ST                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | $\langle \boxed{1}\text{NP}, \boxed{2}\text{VP}[\textit{bse}] \rangle$ |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| <i>word</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                        |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| PHON                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | can                                                                    |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| SUBJ                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | $\langle \boxed{1} \rangle$                                            |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| COMPS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | $\langle \quad \rangle$                                                |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |
| ARG-ST                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | $\langle \boxed{1}\text{NP}, \text{VP}[\textit{pro}] \rangle$          |  |      |     |      |                             |       |                             |        |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |  |      |     |      |                             |       |                         |        |                                                               |

The lexical entry in (19b) will then project the VPE structure (20) for a sentence like (9a):

VP Ellipsis and VP Fronting



In the structure of (20), the head daughter's COMPS list (VP[*pro*]) is elided and is not realized in the syntax. The sentences in (9b) and (9c) are also such cases: verbs such as *has* and *was* are auxiliary verbs ([+AUX]) and subcategorize for a VP complement. Thus, their VP complement can be elided but not that of the main verbs in (10). In the same manner, this analysis will easily generate examples like (21).

- (21) Kim must have been dancing and  $\left. \begin{array}{l} \text{a. Sandy must have been \_\_, too.} \\ \text{b. Sandy must have \_\_, too.} \\ \text{c. Sandy must \_\_, too.} \end{array} \right\}$

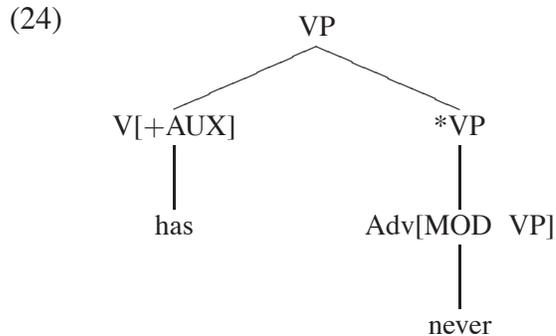
One important constraint on VPE is that it cannot apply immediately after an adverb, as illustrated in (22):

- (22) a. Tom has written a novel, but Peter never has \\_\\_ .  
 b. \*Tom has written a novel, but Peter has never \\_\\_ .

One simple fact we can observe from (22) is that adverbs cannot modify an empty VP. In the framework of HPSG, VP modifying adverbs carry at least the lexical information given in (23).

- (23)  $\left[ \begin{array}{l} \text{HEAD} \left[ \begin{array}{l} \textit{adv} \\ \text{MOD VP:} \boxed{2} \end{array} \right] \\ \text{CONTENT} \left[ \begin{array}{l} \textit{adv-rel} \\ \text{ARG} \quad \boxed{2} \end{array} \right] \end{array} \right]$

The lexical entry in (23) simply states that the adverb with this lexical information modifies a VP. The head feature MOD guarantees the fact that the adverb selects the head VP it modifies. This then entails that when the VP that an adverb modifies is not syntactically realized as in (22b), there is no VP for the adverb to modify. Given Sag and Fodor's (1994) traceless theory,<sup>1</sup> an ungrammatical example like (22b) then would have the structure given in (24).<sup>2</sup>



This explains the unacceptability of VPE after an adverb (cf. Kim 2000).

This analysis also can provide a simple analysis for the puzzling property with respect to the negator *not*: it is possible to strand the negator *not* in VPE when it follows a finite auxiliary, but not when it follows a nonfinite auxiliary verb.

- (25) a. Kim said he could have heard the news, but Lee said that he could not \_ .  
 b. \*Kim said he could have heard the news, but Lee said that he could have not \_ .

Following the analysis of Warner (2000), Kim (2000), and Kim and Sag (2001), I adopt the idea that the English negator *not* leads a double life: one as a nonfinite VP modifier when it is constituent negation and the other as a complement of a finite

<sup>1</sup> See Sag and Fodor (1994) for the critical reviews of positing phonetically empty categories.

<sup>2</sup> But notice that we have a different prediction for modifiers in VP final position:

- (i) a. Tom will not finish his book on Monday, but Kim [[will \_] on Tuesday].  
 b. Kim will not eat the fish with knife, but Kim [[will \_] with chopsticks].

The adverbial elements here are right-adjoined to the VP headed by the auxiliary *will*. This means that there is a VP head daughter the adverbials *on Tuesday* and *with chopsticks* can modify. A similar observation can be found in the contrast between VP final adverbs and VP initial adverbs:

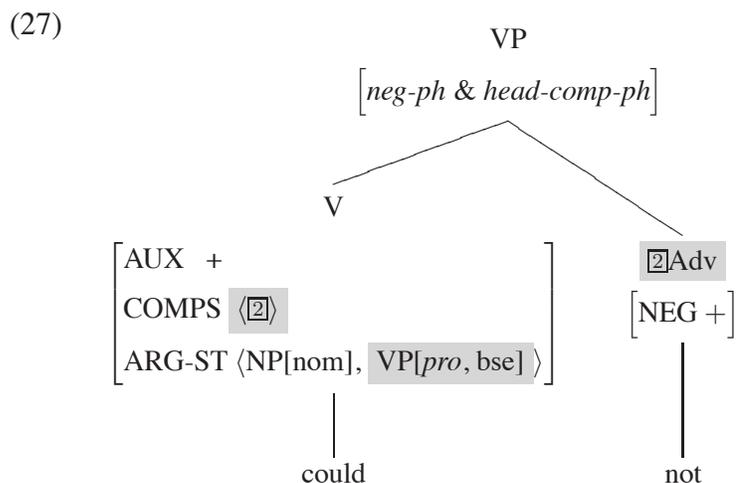
- (ii) a. Kim has been driving the car carelessly, but Mary [[has \_] carefully].  
 b. \*Lee is simply being a student, but Kim is not [simply [\_]].

Though the adverb *carefully* modifies the VP headed by *has*, *simply* has nothing to modify. See Kim (2000) for further discussion.

auxiliary verb when it is sentential negation.<sup>3</sup> Within this analysis, the negator *not* in (25a) (but not the one in (25b)) is the complement of the finite auxiliary verb *could* as given in (26a). But when its VP complement is realized as *pro*, this VP does not appear in the COMPS list as represented in (26b).

- (26) a.  $\left[ \begin{array}{l} \text{COMPS } \langle \boxed{2}, \boxed{3} \rangle \\ \text{ARG-ST } \langle \text{NP}, \boxed{2}\text{ADV}[\text{NEG } +], \boxed{3}\text{VP}[\text{bse}] \rangle \end{array} \right]$
- b.  $\left[ \begin{array}{l} \text{COMPS } \langle \boxed{2} \rangle \\ \text{ARG-ST } \langle \text{NP}, \boxed{2}\text{ADV}[\text{NEG } +], \text{VP}[\text{bse}, \textit{pro}] \rangle \end{array} \right]$

The lexical information (26b) would then project the syntactic structure in (27).



Notice that the phrase [*could not*] in (27) forms a well-formed head-complement structure where *not* is the complement of the head *could*. Nothing blocks this structure.

## 1.2. VPE in Infinitival Clauses

The analysis also immediately predicts the behavior of VPE after the infinitival marker *to*, which has also been taken to be an auxiliary verb, as shown in (28).

- (28) a. Tom wanted to go home, but Peter didn't want to \_\_ .  
 b. Because John persuaded Sally to \_\_ , he didn't have to talk to the reporters.

<sup>3</sup> I assume that this English specific property comes from construction constraints on the phrase *neg(ation)-ph*. This phrase, a subtype of *aux-head-ph*, is peculiar in that it selects a [NEG +] adverbial element as its additional complement. This view of negation is similar to that of Kim (2000) and Kim and Sag (2001).

However, an issue arises from the VPE in rationale clauses or in purpose clauses. As noted by Lobeck (1987) and others, in such clauses the auxiliary infinitive marker *to* does not license VPE:

- (29) a. \*Mag Wildwood came to read Fred's story, and I also came to \_\_ .  
 b. \*John selected Bill to talk to the reporters yesterday, but today he chose Ralph to \_\_ .

The generalization we can draw from such a contrast is that VPE is licensed in an infinitive when the infinitive clause is lexically selected (see Lobeck (1987) for a similar generalization). This could be further supported by the contrast in (30).

- (30) a. Even though he doesn't like to \_\_ , Ron jogs every day.  
 b. \*Even though he could jog to \_\_ , Rob doesn't do anything to stay in shape.

This restriction can be interpreted as a constructional constraint on *ellipsis-ph* in (31).

(31) *ellipsis-ph*

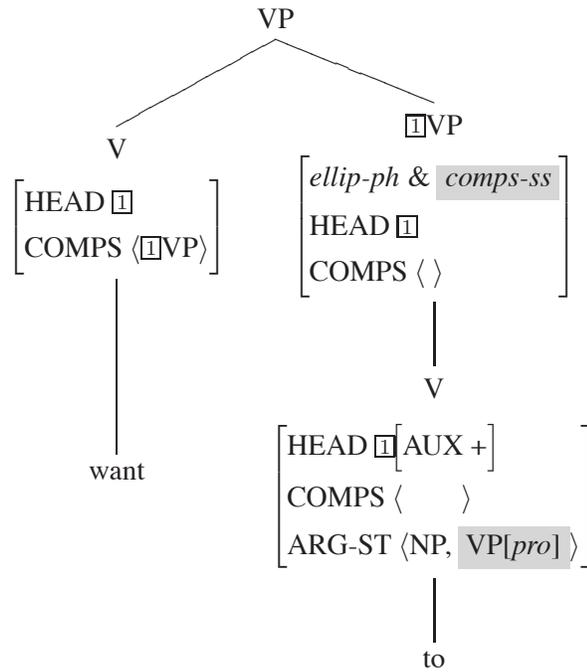
$$\left[ \textit{fin-vp} \vee \textit{comps-ss} \right] \rightarrow \mathbf{H[AUX +]}, \dots$$

What the constraint in (31) ensures is that the *ellipsis-ph* itself should be either a finite clause or an element of the COMPS list. This in turn means that only a complement infinitival VP clause could serve as a VP ellipsis phrase, but not a purpose or rationale clause which is not lexically selected.

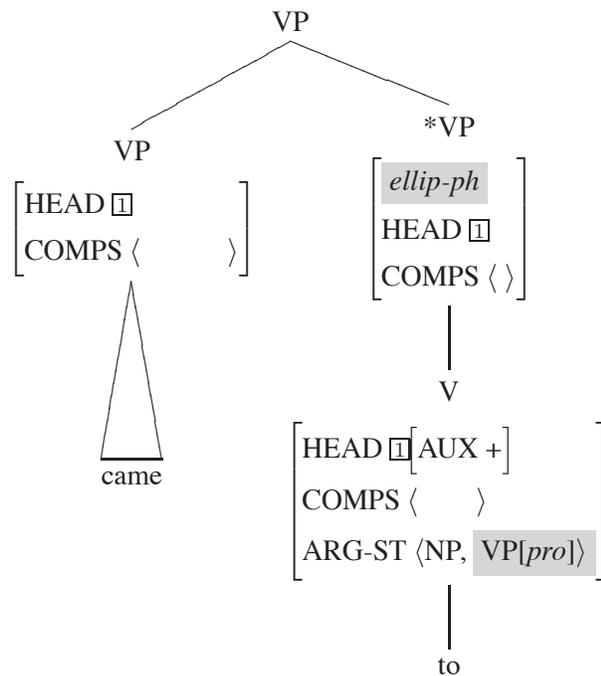
This analysis would generate the structure (32) for (28a) and (33) for the sentence (29a):

VP Ellipsis and VP Fronting

(32)



(33)



As marked by the shaded area in (32) and (33), the marker *to* in both cases selects a base VP as its complement, and hence it can be unrealized in syntax in accordance with the Argument Realization Constraint on auxiliary verbs. But the constraint on *ellipsis-ph* requires the phrase to be *comps-synsem*: Only the VP structure in (32) observes this.

One immediate consequence of this analysis is a clean account of so-called extraposed cases like (34), which have often been treated as a movement process.

- (34) a. You shouldn't talk to reporters because it is dangerous to \_\_ .  
 b. It doesn't bother Mary to tour art galleries, but it certainly annoys Bill to \_\_ .

In a lexicalist perspective, the adjective *dangerous* and the verb *annoys* in such cases will have at least the lexical information given in (35).

- (35) a. *dangerous*: [COMPS ⟨VP[inf]⟩]      b. *annoys*: [COMPS ⟨NP, VP[inf]⟩]

Since the infinitival VP is within the list of COMPS, it could serve as an *ellipsis-ph*. But the ellipsis is blocked in the preverbal infinitival subject position as represented in (36):<sup>4</sup>

- (36) a. \*You shouldn't talk to reporters because [to \_\_] is dangerous.  
 b. \*It doesn't bother Mary to tour art galleries, but [to \_\_] certainly annoys Bill.

The reason for the ungrammaticality of these examples comes from the constructional constraint in (31). No *subject-ss* can serve as an ellipsis phrase. The predicates *dangerous* and *annoys* in (36) would have the lexical information given in (37a) and (37b).

- (37) a. *dangerous* in (36)a:                      b. *annoys* in (36)b:  
           [SUBJ ⟨VP[inf]⟩]                              [SUBJ ⟨VP[inf]⟩]  
           [COMPS ⟨       ⟩]                              [COMPS ⟨NP⟩]

The infinitival VP here is not a *comps-ss* but a *subj-ss* which must be realized in syntax.

---

<sup>4</sup> VPE in infinitives is grammatical when the negative marker *not* precedes *to* in infinitival subjects and adjuncts.

- (i) a. Mary wants to try to get a raise, because [not [to \_\_]] would be silly.  
       b. Mary wants to get a raise, and for her [not [to \_\_]], she'd have to refuse to be evaluated.  
 (ii) a. Mag came to introduce the barkeep but I came [not to \_\_].  
       b. John recounted a story to remember because Bill had recounted a story [not to \_\_].

The data suggest that the infinitival phrase modified by an adverbial element can also serve as an *ellipsis-ph*. But when *not* follows *to*, no VPE is allowed, as our analysis predicts:

- (iii) \*John prefers to talk to Mary, but Bill would like to not \_\_ can help it.

## 2. VP Fronting

Now, let us consider the phenomenon often referred to as VP fronting, given in (38a).

- (38) Mary claimed that
- a. eat bananas, John would \_\_
  - b. \*eating bananas, John started \_\_ .
  - c. \*eat bananas, John made me \_\_ .

The first thing we notice is that VPF cannot be identified with VPE because of ungrammatical cases like those given in (39b) and (39c).

- (39) They swore that Lee might have been using heroin, and
- a. using heroin he might have **been** \_\_ !
  - b. \*been using heroin he might **have** \_\_ !
  - c. \*have been using heroin he **might** \_\_ ! (Akamajian et al. 1979)

If any constituent that can undergo VPE can also be VP-fronted, we would expect that VPs headed by *been* or *have* would not be able to be elided. A simple generalization we can obtain from such data is that a fronted VP cannot be headed by an auxiliary element. This will block examples like (39b) and (39c).

However, this simple approximation is counterevidenced by examples like (40) (data from Gazdar et al. 1982):

- (40) a. \*... and [to go] he is \_\_ .
- b. \*... and [be going] he will \_\_ .
  - c. \*... and [have gone] he will \_\_ .
  - d. ... and [being evasive] he was \_\_ .

Notice here that though the fronted VP in (40a-c) carries the [+AUX] value inherited from the head (*to*, *be* and *have*), they are all unacceptable. To overcome this issue, we assume that the progressive *be* and perfective *have* are aspect verbs ([+ASP(ECT)]) (cf. Gazdar et al. 1982) together with the constructional constraint in (41).

- (41) *vp-filler-ph*

$$\left[ \quad \right] \rightarrow \left[ \begin{array}{l} \textit{filler} \\ \text{ASPECT} \quad - \\ \text{VFORM} \quad \neg \textit{inf} \\ \text{LOCAL} \quad \boxed{1} \end{array} \right] \left[ \begin{array}{l} \text{H} \\ \text{SLASH} \quad \{ \boxed{1} \} \end{array} \right]$$

We thus accept the idea that VPF is different from VPE in that the former is a subtype of a *head-filler-ph* construction constituting a filler and a sentence with this filler as a gap. To allow a VP to be gapped, all we need to do is to slightly revise the Argument Realization Constraint on *auxiliary verbs* as in (42).

(42) Argument Realization Constraint on *aux-verbs* (final):

$$aux\text{-verb} \rightarrow \left[ \begin{array}{l} \text{SUBJ} \quad \boxed{A} \\ \text{COMPS} \quad \boxed{B} \\ \text{ARG-ST} \quad \boxed{A} \oplus \boxed{B} \oplus \text{list}(\text{XP}[\text{noncanon-ss}]) \end{array} \right]$$

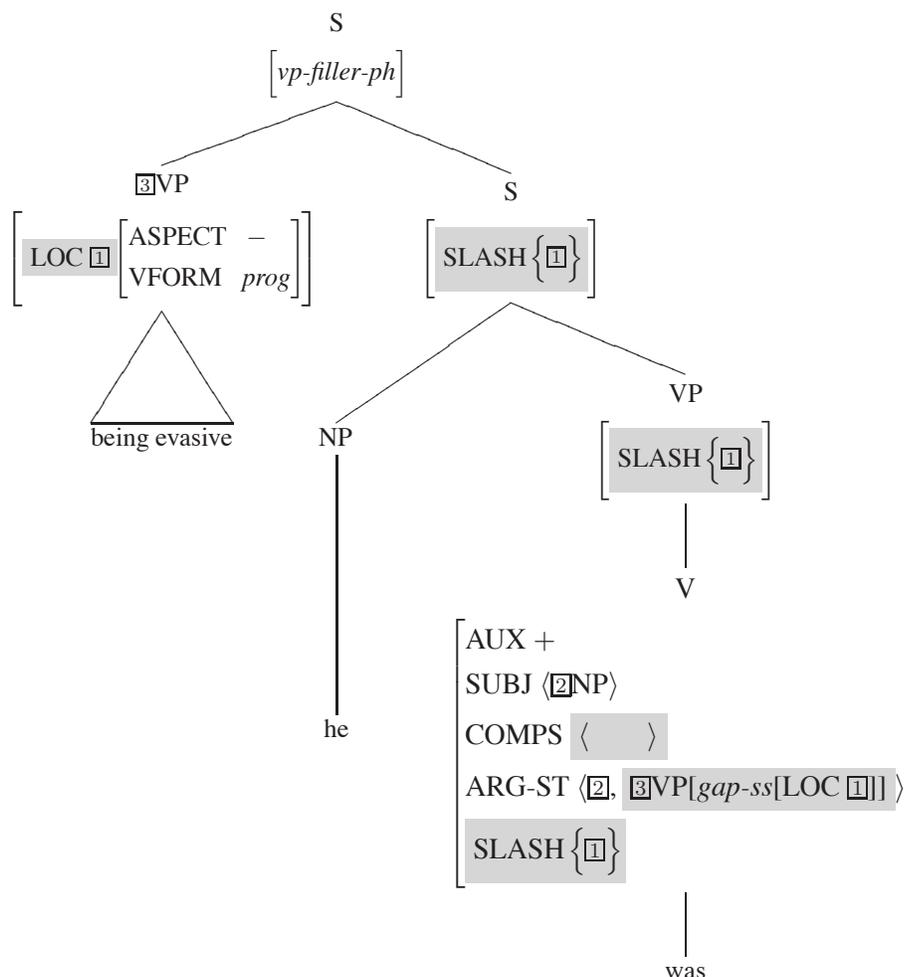
This revised constraint tells that when the final member of the ARG-ST is realized as *noncanon-synsem* whose subtypes are *pro* and *gap-ss*, it does not appear in its COMPS list. Thus when it is realized as *pro*, it introduces *ellipsis-ph* and when it is realized as *gap*, the auxiliary verb lexically introduces a nonempty SLASH value in accordance with the constraint in (43)a. The information in (43)b is partial lexical information for the verb *was* in (40), whose VP complement is topicalized.

(43)

$$\begin{array}{l} \text{a. } [gap\text{-synsem}] \rightarrow \left[ \begin{array}{l} \text{LOC } \boxed{1} \\ \text{SLASH } \{ \boxed{1} \} \end{array} \right] \\ \text{b. } \left[ \begin{array}{l} \text{word} \\ \text{HEAD } [AUX +] \\ \text{SUBJ } \langle \boxed{2} \text{NP} \rangle \\ \text{COMPS } \langle \quad \rangle \\ \text{ARG-ST } \langle \boxed{2}, \text{VP} \left[ \begin{array}{l} \text{gap-ss} \\ \text{LOC } \boxed{1} \end{array} \right] \rangle \\ \text{SLASH } \{ \boxed{1} \} \end{array} \right] \end{array}$$

This lexical information, together with the constraint on *gap-synsem*, would then generate the structure (44) for (40):

(44)



The current system that allows only the complement of an auxiliary verb ([AUX +]) to be realized in a nonlocal position (realized as a *gap-ss*) can easily block overgenerating cases such as those given in (45):

- (45) a. \*I never thought that he would want to go, but [to go] he wanted \_\_ .  
 b. \*I never thought Lee would help move the chair, but [move the chair] Lee helped \_\_ .  
 c. \*I never thought Lee would stop feeding the dog, but [feeding the dog], Lee stopped \_\_ .

The constructional constraint in (41) also imposes constraints on its filler. The slashed VP should be [−ASP], i.e. not headed by progressive *be* or perfective *have*.<sup>5</sup> By requiring such restrictions on the filler VP, we can account for the grammatical contrast in the following examples:

<sup>5</sup> The filler also has the constraint that it should not be infinitive. Such a constraint will block examples like (43a) and (45a).

- (46) a. \*They said he would go, and  $VP_{[+ASP]}$ [be going] he will \_\_ .  
 b. \*They said he would have finished it, and  $VP_{[+ASP]}$ [have finished] he will \_\_ .  
 c. They said he would be noisy, and  $VP_{[-ASP]}$ [being noisy] he was \_\_ .

As assumed, the progressive *be* and perfective *have* in (46a-b) are [+ASP]. The lexical rule specifies that the VP headed by either of these aspect verb cannot undergo the fronting process.

Taking VP fronting to be a special case of topicalization, the analysis also allows unbounded examples like (47) (data from Gazdar et al. 1982).

- (47) a. ... and [go], I think he will \_\_ .  
 b. ... and [going], I believe Kim knew he was \_\_ .  
 c. ... and [being evasive], I believe he knew he was \_\_ .

Notice that this constructional analysis explicitly factors out the similarities and differences between VPF and VPE phenomena. VPF is relevant only to the phrase whose head is an auxiliary. This restriction similarly holds in VPE too. But the restriction on the aspectual value of the gapped VP complement makes them different: the head of the VP that undergoes fronting should be nonaspectual. This accounts for the difference between VPF and VPE (relevant data repeated here).

- (48) They swore that Lee might have been taking heroin, and  
 a. taking heroin he might have been \_\_ !  
 b. \*been taking heroin he might have \_\_ !  
 c. \*have been taking heroin he might \_\_ !
- (49) Lee might have been taking heroin, and  
 a. Sandy might have been \_\_ too.  
 b. Sandy might have \_\_ too.  
 c. (?)Sandy might \_\_ too.

Then, what does this VP fronting analysis predict concerning negation? Recall that our treatment allows *not* to be either a VP modifier or a syntactic complement of a finite auxiliary, and that we permit the negator *not* to be stranded only if it becomes a complement of a finite auxiliary, i.e. only when it occurs immediately after a finite auxiliary. This prediction is borne out:

- (50) a. They all said that John was not being followed, and [being followed] he was not \_\_ .

## *VP Ellipsis and VP Fronting*

- b. They all said that John was not being followed, and [not being followed] he was \_\_ .

The negative marker *not* in (50a) is the complement of the auxiliary *be*, and its VP complement *being followed* is fronted. In (50b), the same VP is fronted and *not* is modifying the fronted VP.

But notice a different behavior of the modifier *not*.

- (51) Kim said she would be not eating spinach, and  
a. \*[eating spinach] she will be not \_\_ .  
b. [not eating spinach] she will be \_\_ .

The negator *not* (51)a can be only a modifier. As noted in the previous section, the modifier *not* cannot be stranded, since the modifier *not* does not satisfy its MOD requirement. Meanwhile, nothing blocks (51)b in which *not* is a modifier.

Within the present analysis, examples like (52) will be predicted to be unacceptable:

- (52) Kim said she would not be eating spinach, and  
a. \*[be eating spinach] she will not \_\_ .  
b. \*[not be eating spinach] she will \_\_ .

The fronted VP is headed by the aspectual head, the progressive *be*. This violates the condition on the assumed *vp-filler-ph*.

### **3. Conclusion**

I have presented an analysis of VP ellipsis and VP fronting within a system in which phrases are modeled as typed feature structures. This theory of grammar, introducing grammatical constructions together with declarative constraints on them, gives us a clean analysis of some otherwise puzzling phenomena in English, VP ellipsis and VP fronting. It has been common practice to accept such phenomena as idiosyncratic in terms of their distributional properties. This has led to the adoption of English-specific rules like *do*-support and the introduction of functional projections such as NegP,  $\Sigma$ P, AgrP, and PredP, as well as transformational operations and empty categories. The construction, constraint-based analysis presented in this paper makes it unnecessary to resort to such abstract machinery, enabling us to find new levels of generalizations within the English auxiliary system, which has often been regarded as a storehouse of peculiarities.

**References**

- Akmajian, Adrian, Susan Steele, and Thomas Wasow. 1979. The Category AUX in Universal Grammar. *Linguistic Inquiry* 10:1–64.
- Bender, Emily M. 2000. *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*. Ph.D. diss., Stanford University.
- Bresnan, Joan. 1982. Control and complementation. *Linguistic Inquiry* 13:343–434.
- Chao, Wynn. 1988. *On Ellipsis*. New York: Garland Publishing, Inc.
- Gazdar, Gerald, Geoffrey K. Pullum, and Ivan Sag. 1982. Auxiliaries and Related Phenomena in a Restrictive Theory of Grammar. *Language* 58:591–638.
- Hardt, Daniel. 1999. VPE as a Proform: Some Consequences for Binding. Ms., Villanova University.
- Johnson, Kyle. 2001. What VP Ellipsis Can Do, and What It Can't, but not Why. In Mark Baltin and Chris Collins, eds., *The Handbook of Contemporary Syntactic Theory*, 439–479. Oxford: Blackwell.
- Kim, Jong-Bok. 2000. *The Grammar of Negation: A Constraint-Based Approach*. Stanford: CSLI Publications.
- Kim, Jong-Bok, and Ivan A. Sag. 2001. Negation with Head Movement. To appear in *Natural Language and Linguistic Theory*.
- Langacker, R. 1966. On Pronominalization and the Chain of Command. In D.A. Reibel and S. Schane, eds., *Modern Studies in English*. New Jersey: Prentice-Hall.
- Lobeck, Anne. 1987. VP Ellipsis in Infinitives. In J. McDonough and B. Plunkett, eds., *NELS*, 425–442.
- Lobeck, Anne. 1995. *Ellipsis: Functional Heads, Licensing and Identification*. Oxford: Oxford University Press.
- López, Luis. 2000. Ellipsis an Discourse-linking. *Lingua* 110:183–213.
- Napoli, D. J. 1985. Verb Phrase Deletion in English: A Base-Generated Analysis. *Journal of Linguistics* 21:281–319.
- Manning, Christopher D. and Ivan Sag. 1999. Dissociations between argument structure and grammatical relations. In G. Webelhuth, J.-P. Koenig, and A. Kathol, eds., *Lexical and Constructional Aspects of Linguistic Explanation*, 63–78. Stanford: CSLI.
- Sag, Ivan A. 1976. *Deletion and Logical Form*. Ph.D. Dissertation, MIT.
- Sag, Ivan A. 2000. Rules and Exceptions in the English Auxiliary System. Ms., Stanford University.
- Sag, Ivan A. and Janet Fodor. 1994. Extraction Without Traces. In Raul Aranovich, William Byrne, Susanne Preuss, and Martha Senturia, eds., *Proceedings of the*

*VP Ellipsis and VP Fronting*

- West Coast Conference on Formal Linguistics 13*, 365–384. Stanford: Stanford Linguistic Association.
- Warner, Anthony R. 2000. English Auxiliaries Without Lexical Rules. In R. Borsley, ed., *Syntax and Semantics, Volume 32: The Nature and Function of Syntactic Categories*, 167–220. San Diego and London: Academic Press.
- Williams, Edwin. 1977. Discourse and Logical Form, *Linguistic Inquiry* 8:101-139.

Jong-Bok Kim  
School of English  
Kyung Hee University  
1 Hoegi-dong, Dongdaemoon-gu, Seoul  
Korea 130-701

[jongbok@khu.ac.kr](mailto:jongbok@khu.ac.kr)



## Relexification: A Process Available to Human Cognition

CLAIRE LEFEBVRE

*Université du Québec à Montréal and University of California, Berkeley*

### 0. Introduction

This paper is about the cognitive process of relexification.<sup>1</sup> Various representations of this process have been proposed in the literature (see e.g. Muysken 1981; Lefebvre and Lumsden 1994a, 1994b; Mous 1995). For the purpose of this paper, I will assume a definition along the lines of Mous' proposal. Given a lexical entry of a language<sub>*i*</sub> of the type in (1a), with a phonological representation, semantic features, and syntactic features, relexification can be seen as the addition of a new phonological representation to this lexical entry. As can be seen in (1b), this new representation is taken from a language<sub>*j*</sub>. Note, however, that the new phonological representation is identified as *j*' rather than as *j* because the new phonological representation is adapted from language<sub>*j*</sub> on the basis of language<sub>*i*</sub>, as is extensively discussed in Lefebvre (1998:16-18) and in Brousseau (in preparation). During a certain period of time, the lexical entry has two phonological representations. In a creole genesis context, eventually, the original phonological representation ceases to be used. The resulting lexical entry has the properties illustrated in (1c)—that is, a phonological representation derived from language<sub>*j*</sub> and semantic and syntactic properties derived from language<sub>*i*</sub>. The process of relexification thus produces new lexical entries that have mixed properties; these properties are mixed in a principled way.

(1) a. Original lexical entry of a language<sub>*i*</sub>

$$\left[ \begin{array}{c} /phonology/_{i} \\ [\text{semantic feature}]_{i} \\ [\text{syntactic feature}]_{i} \\ \text{etc.} \end{array} \right]$$

---

<sup>1</sup> The research underlying this paper has been funded by SSHRCC, FCAR, and FIR-UQAM through several projects that took place at UQAM. I thank the participants to these projects for their input in elucidating the complex problem of creole genesis. I thank Andrée Bélanger for her help in formatting this manuscript.

- b. Addition of a new phonological representation, taken from a language<sub>j</sub>

$$\left[ \begin{array}{l} /phonology/_{i} \quad /phonology/_{j'} \\ [semantic\ feature]_{i} \\ [syntactic\ feature]_{i} \\ etc. \end{array} \right]$$

- c. Loss of the original label resulting in a mixed lexical entry

$$\left[ \begin{array}{l} /phonology/_{j'} \\ [semantic\ feature]_{i} \\ [syntactic\ feature]_{i} \\ etc. \end{array} \right]$$

Relexification constitutes an optimal way of creating a new lexicon in a relatively short period of time. This process has been shown to play a role in the formation of mixed languages such as Media Lengua (see e.g. Muysken 1981), Mechif (see e.g. Bakker 1992), and Ma'a (see e.g. Mous 1995). The global results of 25 years of research at UQAM that I conducted with various colleagues (see e.g. Lefebvre and Kaye 1986, Lefebvre and Lumsden 1989a) show that relexification has also played a major role in the formation of creole languages (see Lefebvre 1998, and the references therein), and that the nature of the process explains the principled division of properties between the contributing languages of a given creole.

Our research consisted of a systematic comparison of the lexicon and grammar of Haitian creole with those of its source languages: French, the superstratum language, and Fongbe, one of the substratum languages. Although there were several languages spoken in Haiti at the time the creole was formed, an in-depth study of one substratum language was preferred over a superficial overview of several of them. The various facets of this methodological choice are extensively discussed in Lefebvre (1998:52-77). Note, however, that, as per the historical research carried out by John Singler for our research (see e.g. Singler 1996), the Gbe languages, Fongbe among them, were predominant in Haiti at the time Haitian creole was formed.

In this paper, I show that the process of relexification explains the principled division of properties of the Haitian lexicon between its contributing languages, French and the West African languages. First, I consider major category lexical entries. Second, I discuss the functional categories involved in the nominal structure. Finally, I examine the consequences of the availability of relexification to human cognition.

### 1. Lexical Semantics

Consider the nouns in (2). The Haitian lexical entries all have two meanings. For example, the noun *plim* means both 'feather' and 'hair'. When we look at the corresponding French lexical entry, we see that the Haitian form is derived from the French; hence, Haitian *plim* is phonologically derived from French *plume*. But

*Relexification: A Process Available to Human Cognition*

the French lexical entry has only one meaning, and therefore it cannot be the source of the extra meaning associated with the Haitian lexical entry. Where does the extra meaning in the Haitian lexical entry come from? When we look at the corresponding Fongbe lexical entry, we see that the latter has the same two meanings as the Haitian one. This suggests that the substratum lexical entry is the source of the semantic properties of the Haitian one. The nature of the process of relexification predicts the data in (2): lexical entries of the type of the Fongbe ones have been relexified on the basis of French, yielding Haitian lexical entries which have phonological representations derived from French and their other properties derived from the substratum languages.

|     |                     |           |                         |
|-----|---------------------|-----------|-------------------------|
| (2) | HAITIAN             | FRENCH    | FONGBE                  |
|     | plim                | plume     | fún                     |
|     | ‘feather’           | ‘feather’ | ‘feather’               |
|     | ‘hair’              |           | ‘hair’                  |
|     | vyann               | viande    | làn                     |
|     | ‘meat’              | ‘meat’    | ‘meat’                  |
|     | ‘edible animals’    |           | ‘edible animals’        |
|     | (complement of      |           | (complement of          |
|     | the verb ‘to kill’) |           | the verb ‘to kill’)     |
|     | dife                | (du) feu  | myòn                    |
|     | ‘fire’              | ‘fire’    | ‘fire’                  |
|     | ‘brand’             |           | ‘brand’                 |
|     | tèt                 | tête      | tà                      |
|     | ‘head’              | ‘head’    | ‘head’                  |
|     | ‘roof’              |           | ‘roof’                  |
|     | van                 | vent      | jòhòn                   |
|     | ‘wind’              | ‘wind’    | ‘wind’                  |
|     | ‘air’               |           | ‘air’                   |
|     |                     |           | (from Lefebvre 1998:71) |

The verbs in (3) illustrate the same phenomena.

Claire Lefebvre

|     |                                  |                                  |                                  |
|-----|----------------------------------|----------------------------------|----------------------------------|
| (3) | HAITIAN                          | FRENCH                           | FONGBE                           |
|     | ansasinèn                        | assassiner                       | hù                               |
|     | ‘to murder’                      | ‘to murder’                      | ‘to murder’                      |
|     | ‘to mutilate’                    |                                  | ‘to mutilate’                    |
|     | gade                             | garder                           | kpón                             |
|     | ‘to watch over/<br>take care of’ | ‘to watch over/<br>take care of’ | ‘to watch over/<br>take care of’ |
|     | ‘to keep’                        | ‘to keep’                        | ‘to keep’                        |
|     | ‘to look’                        | regarder ‘to look’               | ‘to look’                        |
|     | ‘to imitate’                     | imiter ‘to imitate’              | ‘to imitate’                     |
|     | gade                             | regarder                         | kán                              |
|     | ‘to concern’                     | ‘to concern’                     | ‘to concern’                     |
|     |                                  | ‘to look’                        |                                  |
|     | kase                             | casser                           | gbà (-kpó)                       |
|     | ‘to slim down’                   |                                  | ‘to slim down’                   |
|     | ‘to break’                       | ‘to break’                       | ‘to break’                       |
|     | kraze                            | écraser                          | kíjá                             |
|     | ‘to break to pieces’             | ‘to destroy’                     | ‘to break to pieces’             |
|     | ‘to break in<br>spreading’       | ‘to crush’                       | ‘to break in<br>spreading’       |
|     | ‘to disperse’                    |                                  | ‘to disperse’                    |

(from Lefebvre 1999:69-79)

The process of relexification also explains why some French lexical entries have not made their way into Haitian creole. Consider the personal pronouns in (4). Haitian has five personal pronouns with a single entry serving as both first and second person plural. The Haitian forms are phonologically derived from French. But French has six personal pronouns. Why did Haitian end up with only five forms? The answer to this question is predicted by the relexification hypothesis. Fongbe has only five forms with one form serving as both first and second person plural. The creators of Haitian who had a paradigm of personal pronouns of the Fongbe type only had five lexical entries to relexify and so they did. This explains why the extra French form did not make its way into Haitian.

|     |                       |                        |                           |
|-----|-----------------------|------------------------|---------------------------|
| (4) | a. FRENCH             | b. HAITIAN             | c. FONGBE                 |
|     | moi ‘me’              | mwen ‘I, me’           | nyè ‘I, me’               |
|     | toi ‘you (sg.)’       | ou/[wu] ‘you (sg.)’    | hwè ‘you (sg.)’           |
|     | lui/elle ‘he/she/it’  | li ‘he/she/it/him/her’ | é(yè) ‘he/she/it/him/her’ |
|     | nous ‘we/us’          | nou ‘we/us/you (pl.)’  | mí ‘we/us/you (pl.)’      |
|     | vous ‘you (pl.)’      |                        |                           |
|     | eux/elles ‘they/them’ | yo ‘they/them’         | yé ‘they/them’            |

((b) from Valdman et al. 1981, (c)=(18) in Brousseau 1995)

*Relexification: A Process Available to Human Cognition*

The paradigm of wh-words provides a similar example. As can be seen in (5a), French has eight or nine wh-words, depending on whether *que* and *quoi* are considered to be allomorphs or separate lexical entries. The Haitian wh-words in (5b) have phonological representations that are derived from French. But Haitian has only four wh-words. Questions from other positions have to be rendered in wh-phrases, as shown in (5b'). Why have not all the French wh-forms made their way into Haitian, and why did Haitian end up with these exact four lexical entries and not others? Again, the process of relexification provides an answer to both of these questions. As is shown in (5c), the paradigm of Haitian wh-words presents the same inventory of wh-words as the Fongbe one, and both languages express questions from other positions by means of wh-phrases made up of similar elements: a wh-element meaning 'which' and a noun (for a discussion on the difference in morpheme and word order between Haitian and Fongbe, see Lefebvre 1998:171-183, and the references cited therein). The relexification hypothesis also explains why the other French wh-words did not make their way into Haitian: the creators of the creole did not have these extra lexical entries in their lexicon to relexify.

|     |     |                                     |                           |         |
|-----|-----|-------------------------------------|---------------------------|---------|
| (5) | a.  | lequel/laquelle/lesquels/lesquelles | 'which one(s)'            | FRENCH  |
|     |     | qui                                 | 'who'                     |         |
|     |     | que/quoi                            | 'what'                    |         |
|     |     | où                                  | 'where'                   |         |
|     |     | quand                               | 'when'                    |         |
|     |     | comment                             | 'how'                     |         |
|     |     | combien                             | 'how much/how many'       |         |
|     |     | pourquoi                            | 'why'                     |         |
|     | b.  | ki-lès                              | 'which one'               | HAITIAN |
|     |     | (ki-)sa                             | 'what'                    |         |
|     |     | kouman                              | 'how'                     |         |
|     |     | konbyen                             | 'how much, how many'      |         |
|     |     |                                     | (=5) in Brousseau 1995)   |         |
|     | b'. | ki mounn                            | 'which person/who'        | HAITIAN |
|     |     | ki bagay                            | 'which thing/what'        |         |
|     |     | (ki) kote/ki bò                     | 'which place/where'       |         |
|     |     | ki jan                              | 'which manner/how'        |         |
|     |     | ki kalite                           | 'which kind/how'          |         |
|     |     | ki lè                               | 'which time, moment/when' |         |
|     |     | pou ki(-sa)                         | 'for what/why'            |         |

Claire Lefebvre

|             |                 |                           |                          |
|-------------|-----------------|---------------------------|--------------------------|
| c.          | HAITIAN         | FONGBE                    |                          |
| wh-words:   | ki-lès          | dè-té                     | ‘which one’              |
|             | (ki-)sa         | (é-)té/àní                | ‘what’                   |
|             | kouman          | něgbòn                    | ‘how’                    |
|             | konbyen         | nàbí                      | ‘how many/much’          |
| wh-phrases: | ki mounn        | mè té                     | ‘which person/who’       |
|             | ki bagay        | nũ té                     | ‘which thing/what’       |
|             | (ki) kote/ki bò | fí (té)                   | ‘which place/where’      |
|             | ki jan          | àlò té                    | ‘which manner/how’       |
|             | ki kalite       | àlòkpà té                 | ‘which kind/how’         |
|             | ki lè           | hwènù té                  | ‘which moment/time/when’ |
|             | pu ki(-sa)      | (é)té ú(tú)/<br>àní ú(tú) | ‘what cause/why’         |

(from Lefebvre 1998:171-183)

The Haitian compounds in (6) all refer to body parts. The words that are compounded are all phonologically derived from French. But in French, the corresponding body parts are referred to by means of simplexes. The latter did not make their way into Haitian. Why does this situation obtain? Again, this is predicted by the relexification hypothesis. As is shown in (6), Fongbe refers to these same body parts by means of compounds. Again, the creators of Haitian did not have simplexes to relexify in these cases. This explains why the French simplexes did not make their way into Haitian. The data in (6) further show that the creators of Haitian used the principles of word concatenation of their native languages in forming the compounds that we find in Haitian, for the Haitian compounds in (6) are formed on the model of the Fongbe ones (for a discussion on the ordering of words in Haitian and Fongbe compounds, see Lefebvre 1998:339-342, and the references therein).

|     |                  |                            |                       |           |
|-----|------------------|----------------------------|-----------------------|-----------|
| (6) | FRENCH           | HAITIAN                    | FONGBE                |           |
|     | lèvre ‘lip’      | po-bouch ‘skin-mouth’      | nù-fló ‘mouth-skin’   | ‘lip’     |
|     | narine ‘nostril’ | twou-ne ‘hole-nose’        | àntín-dó ‘nose-hole’  | ‘nostril’ |
|     | cil ‘eyelash’    | plim-je ‘hair-eye’         | wùn-dà ‘eye-hair’     | ‘eyelash’ |
|     | nuque ‘nape’     | dèyè-kou ‘back-neck’       | kò-gúdó ‘neck-back’   | ‘nape’    |
|     | crâne ‘skull’    | kalbas-tèt ‘calabash-head’ | tà-ká ‘head-calabash’ | ‘skull’   |
|     |                  | or tèt-kalbas              |                       |           |

(from Brousseau 1989)

In a similar fashion, and as can be seen in (7), where French has simplexes, Haitian has compounds referring to people having certain characteristics. Again, the Haitian compounds are built on the model of the substratum language.

*Relexification: A Process Available to Human Cognition*

|     |                        |                                                |                                    |                        |
|-----|------------------------|------------------------------------------------|------------------------------------|------------------------|
| (7) | FRENCH                 | HAITIAN                                        | FONGBE                             |                        |
| a.  | chauve                 | tèt-chòv<br>(tête-chauve)<br>'head bald'       | tà-súnsún<br>'head-bald'           | 'bald (person)'        |
| b.  | personne-<br>problème  | tèt-chaje<br>(tête-troublé)<br>'head-troubled' | tà-gbà<br>'head-troubled'          | 'problematic (person)' |
| c.  | aveugle                | je-pete<br>(oeil-pété)<br>'eye-burst'          | nùkún-tón-nò<br>'eye-burst-ATT'    | 'blind (person)'       |
| d.  | audacieux              | je-chèch<br>(oeil-sec)<br>'eye-dry'            | hǎn-wùn<br>'clear-eye'             | 'audacious (person)'   |
| e.  | prétentieux            | je-fò<br>(oeil-fort)<br>'eye-strong'           | nùkún-kèn<br>'eye-strong'          | 'pretentious (person)' |
| f.  | entêté<br>or tête-dure | tèt-di<br>(tête-dure)<br>'head-hard'           | tà-mè-sièn-tó<br>'head-in-hard-AG' | 'stubborn (person)'    |

(from Lefebvre 1998:336)

The distribution of light verb constructions versus simplexes in the three languages under comparison constitutes yet another example of relexification. As is shown by the examples in (8), Haitian has a series of light verb constructions. The Haitian examples parallel the Fongbe ones. The Haitian and Fongbe data contrast with French which uses simplexes to refer to the same notions. Again, the French simplexes in (8) did not make their way into Haitian because the creators of Haitian did not have corresponding simplexes to relexify. And again, the concatenation of light verbs and their objects in Haitian follows the pattern of the substratum language.

|     |    |           |                      |         |
|-----|----|-----------|----------------------|---------|
| (8) | a. | bat       | men                  | HAITIAN |
|     |    | xò        | àsí                  | FONGBE  |
|     |    | hit       | hand                 |         |
|     |    |           | 'to applaud/to clap' |         |
|     |    | applaudir |                      | FRENCH  |
|     |    |           | 'to applaud/to clap' |         |

|    |          |                          |         |
|----|----------|--------------------------|---------|
| b. | bay      | chenn                    | HAITIAN |
|    | ná       | kàn                      | FONGBE  |
|    | give     | chain                    |         |
|    |          | ‘to wind (e.g. a clock)’ |         |
|    | remonter |                          | FRENCH  |
|    |          | ‘to wind’                |         |
| c. | manje    | kòb                      | HAITIAN |
|    | dú       | àkwé                     | FONGBE  |
|    | eat      | money                    |         |
|    |          | ‘to spend’               |         |
|    | dépenser |                          | FRENCH  |
|    |          | ‘to spend’               |         |
| d. | mete     | digo                     | HAITIAN |
|    | dó       | àhò                      | FONGBE  |
|    | put      | indigo                   |         |
|    |          | ‘to tie-dye’             |         |
|    | teindre  |                          | FRENCH  |
|    |          | ‘to tie-dye’             |         |

(from Lefebvre 1999:80-81)

The examples in (2)-(8) all illustrate the fact that relexification has played a role in the formation of Haitian creole. Having examined various cases involving major category lexical entries, I now turn to the discussion of the functional categories involved in the nominal structure.

## 2. Functional Categories Involved in the Nominal Structure

The data in (9) provide an overview of the French nominal structure. They show that in this language the definite determiner, the possessive, and the demonstrative determiners all precede the head noun, and that there can be only one of these per noun phrase. Singular and plural forms are contrasted in (9b) showing that plural is encoded in a bound morpheme in French.

|     |    |      |         |     |          |        |
|-----|----|------|---------|-----|----------|--------|
| (9) | a. | * le | mon     | ce  | crabe    | FRENCH |
|     |    | DET  | POSS    | DEM | crab     |        |
|     | b. | {    | le/les  | }   | crabe    | FRENCH |
|     |    | {    | mon/mes | }   |          |        |
|     |    | {    | ce/ces  | }   |          |        |
|     |    | {    | the     | }   | crab(s)’ |        |
|     |    | {    | my      | }   |          |        |
|     |    | {    | this    | }   |          |        |

(from Lefebvre 1998:78)

*Relexification: A Process Available to Human Cognition*

The Haitian and Fongbe nominal structures are illustrated in (10). In both languages, the determiners all follow the head noun. In both languages, a possessor phrase, a demonstrative term, the definite determiner, and the plural marker may all co-occur within the same nominal structure. In both languages, the plural marker is an independent morpheme.

- |      |                                                           |           |     |     |     |         |
|------|-----------------------------------------------------------|-----------|-----|-----|-----|---------|
| (10) | krab                                                      | [mwɛn ø]  | sa  | a   | yo  | HAITIAN |
|      | àsón                                                      | [nyè tòn] | éló | ó   | lê  | FONGBE  |
|      | crab                                                      | me        | GEN | DEM | DET | PL      |
|      | ‘these/those crabs of mine (in question/that we know of)’ |           |     |     |     |         |
|      | (from Lefebvre 1998:78)                                   |           |     |     |     |         |

The Haitian and Fongbe nominal structures thus contrast in the same way with the French nominal structure with respect to word order, co-occurrence restrictions of determiners, and with respect to whether the plural marker is a free (in Haitian and Fongbe) or a bound (in French) morpheme.

Moreover, with the exception of their phonological representation, the properties of the definite determiners are the same in Haitian and in Fongbe; these properties contrast in a systematic way with those of the French definite determiner. These contrastive properties, discussed at length in Lefebvre (1998:79-84), are summarized in (11).

- |      |                                         |                                               |
|------|-----------------------------------------|-----------------------------------------------|
| (11) | FRENCH [+def.] determiner               | HAITIAN/FONGBE [+def.] determiner             |
|      | – pre-nominal                           | – post-nominal                                |
|      | – marked for gender and number          | – unmarked for gender and number              |
|      | – allomorphs: <i>le/la/les/l’</i>       | – allomorphs: <i>la, a, an, nan, lan/ó,ón</i> |
|      | – anaphoric and cataphoric              | – anaphoric                                   |
|      | – partitive <i>du/des</i>               | – no partitive                                |
|      | – obligatory with generic or mass nouns | – impossible with generic or mass nouns       |
|      | – no bare NPs                           | – bare NPs                                    |
|      | – *Det [relative clause] N              | – N [relative clause] Det                     |

Furthermore, the definite determiners involved in the Haitian and Fongbe nominal structures also play a crucial role in the clause structure of these two languages, as is exemplified in (12) (for an extensive discussion of these facts, see Lefebvre 1998:219-247).

- |      |                                                     |    |      |   |         |
|------|-----------------------------------------------------|----|------|---|---------|
| (12) | a.                                                  | Li | rive | a | HAITIAN |
|      | b.                                                  | É  | wá   | ó | FONGBE  |
|      | ‘He has arrived (as expected/as we knew he would).’ |    |      |   |         |

A Haitian or Fongbe nominal structure may contain a noun followed by the plural marker only, as is shown in (13). In such a case, the structure is interpreted as definite. Comparable data are impossible in French.

- |      |      |                |  |         |
|------|------|----------------|--|---------|
| (13) | krab | yo             |  | HAITIAN |
|      | àsón | lê             |  | FONGBE  |
|      | crab | PL             |  |         |
|      |      | 'the crabs'    |  |         |
|      | *    | '(some) crabs' |  |         |
- (=(31) in Lefebvre 1994)

The data in (14) show that Haitian and Fongbe both allow for bare NPs. Bare NPs are not allowed in French.

- |      |    |           |                 |         |
|------|----|-----------|-----------------|---------|
| (14) | M' | achte     | krab.           | HAITIAN |
|      | N' | xò        | àsón.           | FONGBE  |
|      | I  | buy       | crab            |         |
|      |    | 'I bought | '(some) crabs.' |         |
- (=(32) in Lefebvre 1994)

The data in (15) show that in both Haitian and Fongbe, when the definite determiner and the plural marker co-occur within the same nominal structure, the definite determiner must precede the plural marker.

- |      |      |     |            |                |   |    |   |         |
|------|------|-----|------------|----------------|---|----|---|---------|
| (15) | krab | la  | yo         | /              | * | yo | a | HAITIAN |
|      | àsón | ó   | lê         | /              | * | lê | ó | FONGBE  |
|      | crab | DET | PL         |                |   |    |   |         |
|      |      |     | 'the crabs | (in question)' |   |    |   |         |
- (=(33) in Lefebvre 1994)

Finally, in both languages, there is variation between speakers with respect to the possibility of co-occurrence of the determiner and of the plural marker. Crucially, the patterns of variation are the same in both languages. Two slightly different grammars have been reported on in the literature. They are summarized in (16).

*Relexification: A Process Available to Human Cognition*

- |                |                                                                                                                                                                                                                               |                                                                                                 |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| (16)           | HAITIAN                                                                                                                                                                                                                       | FONGBE                                                                                          |
| G <sub>1</sub> | where <i>la</i> and <i>yo</i> can co-occur<br>(d'Ans 1968:105, Faine 1937:83, Fournier 1977:43, Goodman 1964:45, Joseph 1988:201, Lefebvre and Massam 1988:215, Ritter 1992:207-209, Sylvain 1936:55, Valdman 1978:1994-1995) | where <i>jó</i> and <i>lé</i> can co-occur<br>(Brousseau and Lumsden 1992:22, Lefebvre 1998:85) |
| G <sub>2</sub> | where <i>la</i> and <i>yo</i> cannot co-occur<br>(DeGraff 1992:107, Joseph 1988:201, Lumsden 1989:65)                                                                                                                         | where <i>jó</i> and <i>lé</i> cannot co-occur<br>(Agbidinoukoun 1991:149)                       |

The French and Haitian paradigms of deictic terms are also strikingly different, whereas the Haitian and Fongbe paradigms of deictic terms are strikingly parallel. Due to space limitations, suffice to say here that, while French has eleven deictic terms that can be involved in the nominal structure, Haitian and Fongbe have two. These are shown in (17).

- |      |         |        |
|------|---------|--------|
| (17) | HAITIAN | FONGBE |
|      | sa      | (é)lɔ́ |
|      | sila    | (é)né  |

In Lefebvre (1997, 1998:89-101), it is extensively argued that the properties of the two Haitian terms are not the same as those of the French lexical entries (*ça* and *celui-là*, respectively) which were the source of the phonological representation of the Haitian ones; it is further extensively argued that the two Haitian terms do have the same distributional and syntactic properties as the Fongbe corresponding ones. Furthermore, in Lefebvre (in press), it is shown that in both Haitian and Fongbe, there are three interpretive patterns (identified below as G1, G2 and G3) for the pairs of deictic terms. These are shown in (18) and (19), respectively. Crucially, these patterns are identical for Haitian and Fongbe.

- |      |                |                  |                    |         |
|------|----------------|------------------|--------------------|---------|
| (18) | G <sub>1</sub> | sa [+ proximate] | sila [– proximate] | HAITIAN |
|      | G <sub>2</sub> | sa [α proximate] | sila [– proximate] |         |
|      | G <sub>3</sub> | sa [α proximate] | sila [α proximate] |         |
- Sources: G1: Tinelli (1970:28), Goodman (1964:51)  
 G2: Lefebvre (1997) [see also data in Sylvain (1936) and in Étienne (1974)]  
 G3: Férère (1974:103), Valdman (1978:194), Valdman et al. (1981), Joseph (1989), and my own field notes

- (19) G1 (é)lɔ [+ proximate] (é)né [- proximate] FONGBE  
 G2 (é)lɔ [α proximate] (é)né [- proximate]  
 G3 (é)lɔ [α proximate] (é)né [α proximate]  
 Sources: G1: Anonymous (1983), Segurola (1963), and my own field notes  
 G2: Lefebvre (1997)  
 G3: My own field notes

The data discussed in this section, (9)-(19), show the remarkable parallel that exists between the nominal structures of Haitian and Fongbe. As is extensively argued in Lefebvre (1998:89-101, in press), the extraordinary similarity that exists between the functional categories of the Haitian and Fongbe nominal structures follows from relexification.

### 3. Theoretical Consequences

On the basis of a small set of data, involving minor as well as major syntactic category lexical items, I have shown that the process of relexification explains in a straightforward way the mixed properties of the Haitian lexicon. A more exhaustive comparison of the three lexicons under study may be found in Lefebvre (1998, 1999, in press, and the references therein). My conclusion is that the process of relexification has played a major role in the formation of Haitian creole.<sup>2</sup> I extend this conclusion to other creoles as well. I assume, however, that the quantity of relexified lexical entries may vary between creoles due to variable factors such as the amount of exposure to the superstratum language, etc.

The fact that relexification can be argued to have been at work in various cases of mixed languages and of creole languages argues that this mental process is available to human cognition. It is used as a means of creating new languages in a relatively short time. The fact that it exists and the very nature of the process support Sproat's (1985) and Pranka's (1983) proposal that phonological representations are stored independently in the brain. By definition, the process is used by speakers who already master the properties of a mature lexicon, hence by adults. This challenges the claim that adults have a marginal role in linguistic change (for a discussion of this point, see also Hopper and Traugott 1993:21). The process creates hybrid languages straying from the normal course of linguistic change. This poses a problem for the genetic classification of the languages so formed. If it is the case that relexification is a major process in creole genesis, it calls into question the assumption that all creole languages are alike, as is advocated in Bickerton (1981, 1984). Indeed, due to the very nature of the process, creole languages reproduce the semantic and syntactic properties of their substratum languages.

<sup>2</sup> For a thorough discussion on the complexity of the linguistic situations in which relexification applies in creole genesis, showing that Haitian is not simply a relexified version of Fongbe, see Lefebvre 1998:15-50, in press). For a discussion of the other processes involved in creole genesis, see the same sources and the references cited therein.

## References

- Agbidinougoun, Cosme Christian. 1991. Analyse contrastive des syntagmes nominaux du fongbe et du français. Ph.D. dissertation, Université Paris III (Sorbonne Nouvelle), Paris.
- Anonymous. 1983. *Éléments de recherche sur la langue Fon*. Cotonou.
- Ans, André-Marcel d'. 1968. *Le Créole français d'Haïti: Étude des unités d'articulation, d'expansion et de communication*. The Hague: Mouton.
- Bakker, Peter. 1992. A Language of Our Own: The Genesis of Mitchif, the Mixed Cree-French Language of the Canadian Metis. Ph.D. dissertation, University of Amsterdam, Amsterdam.
- Bickerton, Derek. 1981. *Roots of Language*. Ann Arbor, MI: Karoma.
- Bickerton, Derek. 1984. The Language Bioprogram Hypothesis. *The Behavioral and Brain Sciences* 7:173-221.
- Brousseau, Anne-Marie. 1989. De 'nù-fló' à 'po-bouche': hypothèses sur l'origine des composés en haïtien. In C. Lefebvre and J. S. Lumsden, eds., 285-312.
- Brousseau, Anne-Marie. 1995. Les pronoms en créole haïtien, en français et en fongbé. In C. Lefebvre, ed., volume IV, 33 pages.
- Brousseau, Anne-Marie. In preparation. *The Genesis of the Phonological System of Haitian Creole* [provisional title].
- Brousseau, Anne-Marie, and John S. Lumsden. 1992. Nominal Structure in Fongbe. In C. Lefebvre, ed., 5-26.
- DeGraff, Michel A. F. 1992. Creole Grammars and Acquisition of Syntax: The Case of Haitian. Ph.D. dissertation, University of Pennsylvania, Philadelphia.
- Étienne, Gérard. 1974. Le créole du nord d'Haïti: Étude des niveaux de structure. Ph.D. dissertation, Université des sciences humaines de Strasbourg, Strasbourg.
- Faine, Jules. 1937. *Philologie créole: études historiques et étymologiques sur la langue créole d'Haïti*. Port-au-Prince: Imprimerie de l'État.
- Férère, Gérard A. 1974. Haitian Creole Sound-System, Form-Classes, Texts. Ph.D. dissertation, University of Pennsylvania, Philadelphia.
- Fournier, Robert. 1977. N ap fè yun ti-kose su la (La grammaire de la particule *la* en créole haïtien). Master's thesis, Université du Québec à Montréal, Montreal.
- Goodman, Morris F. 1964. *A Comparative Study of Creole French Dialects*. The Hague: Mouton.
- Highfield, Arthur R., and Albert Valdman, eds. 1981. *Historicity and Variation in Creole Studies*. Ann Arbor, MI: Karoma.
- Hopper, Paul J., and Elizabeth C. Traugott. 1993. *Grammaticalisation* [Cambridge Textbooks in Linguistics]. Cambridge: Cambridge University Press.
- Joseph, Frantz. 1989. La détermination nominale en créole haïtien. Ph.D. dissertation, Université de Paris VII, Paris.
- Lefebvre, Claire, ed. 1992. *Topics on the Syntax and Semantics of Fongbe*, special issue of *Journal of West African Languages* 22.

Claire Lefebvre

- Lefebvre, Claire. 1994. Functional Categories and Relexification. In C. Lefebvre and J. S. Lumsden, eds., 74 pages.
- Lefebvre, Claire. 1997. Relexification in Creole Genesis: The Case of Demonstrative Terms in Haitian Creole. *Journal of Pidgin and Creole Languages* 12:181–201.
- Lefebvre, Claire. 1998. *Creole Genesis and the Acquisition of Grammar: The Case of Haitian Creole* [Cambridge Studies in Linguistics 88]. Cambridge: Cambridge University Press.
- Lefebvre, Claire. 1999. Substratum Semantics in the Verbal Lexicon of Haitian Creole. *Studies in Language* 23:61–103.
- Lefebvre, Claire. In press. The Interplay of Relexification and Levelling in Creole Genesis and Development. *Linguistics*.
- Lefebvre, Claire, and Jonathan Kaye, eds. 1986. *Études syntaxiques, morphologiques et phonologiques*. Research report prepared for FCAR and PAFAC on the project Fon-Créole Haïtien, Université du Québec à Montréal, Montreal.
- Lefebvre, Claire, and John S. Lumsden. 1989a. Les langues créoles et la théorie linguistique. In C. Lefebvre and J. S. Lumsden, eds., 249–272.
- Lefebvre, Claire, and John S. Lumsden, eds. 1989b. *La Créolisation*, special issue of *Revue canadienne de linguistique* 34(3).
- Lefebvre, Claire, and John S. Lumsden. 1994a. Le rôle central de la relexification dans la genèse des langues créoles. *Plurilinguismes* 8:47–93.
- Lefebvre, Claire, and John S. Lumsden. 1994b. Relexification in Creole Genesis. In C. Lefebvre and J. S. Lumsden, eds., 28 pages.
- Lefebvre, Claire, and John S. Lumsden, eds. 1994c. *The Central Role of Relexification in Creole Genesis: The Case of Haitian Creole*, research report prepared for SSHRCC on the project La Genèse du créole haïtien: Un cas particulier d'investigation sur la forme de la grammaire universelle (1 volume), Université du Québec à Montréal, Montreal.
- Lefebvre, Claire, John S. Lumsden, and Lisa Travis, eds. 1992. *Functional Categories*. Special issue of the *Revue canadienne de linguistique* 37.2.
- Lefebvre, Claire, and Diane Massam. 1988. Haitian Creole Syntax: A Case for DET as Head. *Journal of Pidgin and Creole Languages* 3:213–243.
- Lumsden, John S. 1989. On the Distribution of Determiners in Haitian Creole. In C. Lefebvre and J. S. Lumsden, eds., 64–93.
- Mous, Maarten. 1995. Language Intertwining. Paper read at the Amsterdam Creole Workshop: Creole Genesis and Language Contact, Amsterdam.
- Muysken, Pieter C. 1981. Half-Way between Quechua and Spanish: The Case for Relexification. In A. R. Highfield and A. Valdman, eds., 52–79.
- Pranka, Paula M. 1983. Syntax and Word Formation. Ph.D. dissertation, MIT, Cambridge, MA.
- Ritter, Elizabeth. 1992. Cross-Linguistic Evidence for Number Phrase. In C. Lefebvre, J. S. Lumsden, and L. Travis, eds., 197–219.
- Seguro, R.P.B. 1963. *Dictionnaire fon-français*. Cotonou: Procure de l'archidiocèse.

*Relexification: A Process Available to Human Cognition*

- Singler, John V. 1996. Theories of Creole Genesis, Sociohistorical Considerations, and the Evaluation of Evidence: The Case of Haitian Creole and the Relexification Hypothesis. *Journal of Pidgin and Creole languages* 11:185-231.
- Sproat, Richard W. 1985. On Deriving the Lexicon. Ph.D. dissertation, MIT, Cambridge, MA.
- Sylvain, Suzanne. 1936. *Le créole haïtien: Morphologie et syntaxe*. Wetteren, Belgium: Imprimerie De Meester/Port-au-Prince: By the author.
- Tinelli, Henri. 1970. Generative Phonology of Haitian Creole. Ph.D. dissertation, University of Michigan, Ann Arbor.
- Valdman, Albert. 1978. *Le créole: structure, statut et origine*. Paris: Klincksieck.
- Valdman, Albert et al. 1981. Haitian Creole-English-French Dictionary [2 volumes]. Bloomington: Indiana University, Creole Institute.

Claire Lefebvre  
Département de linguistique  
Université du Québec à Montréal  
CP 8888 succ.  
Centre-ville, Montreal  
Canada H3P 3P8

lefebvre.claire@uqam.ca



## Contact, Attrition, and Structural Shift: The Case of Oroqen \*

FENGXIANG LI

*California State University, Chico*

### 0. Introduction

Oroqen is a Tungusic language spoken by roughly 2,500 people in two provinces of northeastern China: Heilongjiang Province and Inner Mongolia Autonomous Region. The Oroqen lived as hunter-gatherers until the early 1950s, when the Chinese government began its effort to settle them. Before this time very little was known about the Oroqen language outside of piecemeal information provided by the great Russian ethnographer, S. M. Shirokogorov (1923, 1929-1933, 1935, 1944). During the early 1960s, after the demise of Oroqen nomadicism, more comprehensive research on the language began with the pioneering work of Hu Zengyi, resulting in the first grammatical sketch of Oroqen (ultimately published as Hu 1986). One additional grammatical sketch has since appeared (Zhang et al. 1989) as well as two phrasebooks (Han and Meng 1993, Saxirong 1981), a collection of texts (Meng 1993), and work on more specific aspects of the grammar (Doerfer 1983; Li and Whaley 1998; B. Li 1992, 1996; F. Li 1996; S. Li 1981; Whaley and Li 1998, 2000; Whaley et al. 1999; Zhang 1996; Li and Whaley 2000). It is glaringly obvious among these works that hardly any attention has been devoted to the rapid changes that Oroqen has been undergoing over the last few decades. Therefore, the aim of this paper is to explore both the internal and external factors that led to the drastic changes that have taken place in Oroqen, focusing on the dismantling and reordering of its structural system. The first section of this paper is a brief introduction. The second section is devoted to a discussion of the changes that are attributable to language attrition. The third section deals with contact-induced changes, and the final section of the paper offers some explanations and systematic patterns about contact and attrition situations.

---

\* This project was partially funded by NSF grant SBR9710091. I would like to thank Juha Janhunen, who read my paper and gave very useful comments. I would also like to thank Lindsay J. Whaley, who did fieldwork with me on Oroqen over the past five years. I am indebted to all of the informants for their generous contributions of data. I benefited from useful discussions with Lenore A. Grenoble and Graham Thurgood. Needless to say, all mistakes remain my own.

### 1. Attrition-Induced Changes

It has been abundantly demonstrated in the literature that the structure of languages undergoing attrition is subject to a certain amount of change at all levels: phonological, morphological, syntactic, semantic, and lexical (Hill 1980, Dorian 1989, Maher 1991, Grenoble and Whaley 1998). The changes that have been described in the literature display systematic patterns. As Schmidt (1985) and Maher (1991) point out, the patterns found in obsolescing languages are remarkably similar—with many showing allomorphic reduction, radical simplification and regularization of certain paradigms, generalization of a single case affix to cover various peripheral case functions, a tendency to eliminate verbal inflectional affixes, a breakdown in agreement rules, replacement of synthetic forms by analytic ones or by periphrastic constructions, and so on. Fieldwork on Oroqen done by Lindsay J. Whaley and Fengxiang Li over the past five years yielded data consistent with some of the aforementioned characteristic features.

For instance, morphological processes that have restricted applications are no longer present in younger Oroqen speakers' speech. Li and Whaley (2000) give a detailed description of one of the intensive marking strategies utilized in Oroqen—namely, emphatic reduplication. Emphatic reduplication operates in Oroqen by copying the first syllable of an adjectival stem onto a CVC template and prefixing the copied material to the adjective. If the first syllable of the adjective is open, then a [b] is inserted into the post-vocalic slot of the CVC template. These patterns are demonstrated in (1).

|     |          |                 |              |                              |
|-----|----------|-----------------|--------------|------------------------------|
| (1) | bagdarin | 'white'         | bag-bagdarin | 'very white, white as snow'  |
|     | ʃɪŋarin  | 'yellow'        | ʃib-ʃɪŋarin  | 'very yellow, golden yellow' |
|     | kara     | 'dark (glossy)' | kab-kara     | 'very dark, glossy black'    |
|     | kəŋɔrin  | 'black'         | kəb-kəŋɔrin  | 'very black'                 |

For certain speakers, the post-vocalic [b] in the last three forms in (1) has assimilated in voicing to the following obstruent, and so is pronounced [p].

Emphatic reduplication occurs only with a small number of adjective stems, perhaps only the four presented in (1). Notably, all the adjectives are color terms, and all of them are commonly associated with the hue of certain domestic animals. Speakers routinely reject emphatic reduplication with other adjectives, even if they are color terms, as shown in (2).

|     |                       |             |                             |
|-----|-----------------------|-------------|-----------------------------|
| (2) | gɪlbarin              | 'sky blue'  | *gɪl-gɪlbarin               |
|     | tʃəŋg <sup>y</sup> ɛn | 'deep blue' | *tʃəŋ-tʃəŋg <sup>y</sup> ɛn |
|     | Ula:rin               | 'red'       | *Ub-Ula:rin                 |
|     | tʃuturin              | 'green'     | *tʃub-tʃuturin              |

*Contact, Attrition, and Structural Shift in Oroqen*

The restrictive application of reduplication has made it highly vulnerable to rapid loss as Oroqen becomes obsolescent. Data from four informants demonstrate that younger speakers no longer have emphatic reduplication in their grammar. Our two older informants A (70 years old) and B (60 years old) both used Oroqen as their primary language until their early twenties. Since that time they have increasingly used Mandarin Chinese. Both women use the reduplicated adjectives in (1), though the younger of the two does not recognize or accept the last of these forms. Our two younger informants C (46 years old) and D (30 years old) no longer accept or recognize any of these forms. Though both of these women learned Oroqen in the home as a first language, both have grown up and been educated in a Mandarin-dominated context.

Similar patterns were found with other morphological processes indicating the same process of loss. Among the nominalizers in Oroqen, *-ŋki* is by far the most productive. It occurs with practically any verb to yield a noun. Examples are given in (3).

- (3)
- |                  |                        |
|------------------|------------------------|
| <i>kadı-rə-n</i> | ‘cut-future-3sg.’      |
| <i>kadı-ŋki</i>  | ‘sickle/scythe’        |
| <i>dʒik-tə-n</i> | ‘slice-nonfuture-3sg.’ |
| <i>dʒiki-ŋki</i> | ‘cutting board’        |
| <i>tək-tə-n</i>  | ‘sit-nonfuture-3sg.’   |
| <i>təyə-ŋki</i>  | ‘sth. to sit on’       |

Both our older informants A and B readily produced many examples using this and other less productive nominalizers. For example, they employed the nominalizer *-wun* (*kadı-wun* ‘knife’, *tukti-wun* ‘ladder’) when they produced the forms for ‘knife’ and ‘ladder’.

What is noteworthy is that quite a few of the less productive nominalizers are no longer in the repertoire of our informants C and D. Our informant C only used the suffix *-ŋki*, even in forms where older speakers do not use it. Although she retains this most productive nominalizer *-ŋki*, and in fact uses it more generally than older speakers, she alternates between having the velar nasal and leaving it out. A similar pattern held for informant D’s speech. Only the most productive nominalizer was employed. However, for her, nasal deletion in the suffix *-ŋki* is compulsory, which is shown in the examples in (4).

- (4)
- |                  |                               |
|------------------|-------------------------------|
| <i>ləxə-rə-n</i> | ‘hang sth. up-nonfuture-3sg.’ |
| <i>ləxə-ki</i>   | ‘hook’                        |
| <i>təyə-rə-n</i> | ‘sit-nonfuture-3sg.’          |
| <i>təyə-ki</i>   | ‘sth. to sit on’              |

|            |                              |
|------------|------------------------------|
| mu:lə:     | ‘water-LOC’                  |
| mu:lə:-ki  | ‘water bucket’               |
| tukti-rə-n | ‘go up/climb-nonfuture-3sg.’ |
| tukti-ki   | ‘ladder’                     |

A clear pattern emerges in the data from the four informants A, B, C, and D. The older speakers have a much richer repertoire of derivational morphology than the younger ones, and there is a gradual phonological erosion of the productive suffixal nominalizer *-ŋki* that correlates with the age of the speakers, which is illustrated in the table in (5).

|     |            |            |                      |           |                  |
|-----|------------|------------|----------------------|-----------|------------------|
| (5) | A          | B          | C                    | D         |                  |
|     | təyə-ŋki   | təyə-ŋki   | təyə-ŋki/təyə-ki     | təyə-ki   | ‘sth. to sit on’ |
|     | mu:lə:-ŋki | mu:lə:-ŋki | mu:lə:-ŋki/mu:lə:-ki | mu:lə:-ki | ‘water bucket’   |
|     | tukti-wun  | tukti-wun  | tukti-ŋki/tukti-ki   | tukti-ki  | ‘ladder’         |
|     | kadı-wun   | kadı-wun   | kadı-ŋki/kadı-ki     | kadı-ki   | ‘sickle/scythe’  |

These data demonstrate, for the younger generation of Oroqen speakers, the wholesale loss of certain limited derivational strategies such as emphatic reduplication, and the replacement of low productivity derivational morphemes such as *-wun* by equivalent, yet more productive, strategies. These changes are fairly consistent with the findings in the literature on first language attrition. Maher (1991:68) points out that indigenous languages undergoing attrition due to the encroachment of a dominant language reveal a restructuring or reconfiguration of morphological and syntactic structures which display the following characteristics: (a) reduction in the number of allomorphs (i.e. more invariable forms, or fewer context sensitive rules) with increased paradigmatic regularity, (b) replacement of synthetic forms by analytic ones or by periphrastic constructions, (c) progressive reduction in inflectional morphology, entailing less flexible word order, (d) preference for coordinate rather than embedded constructions, and (e) distinctive aspectual constructions in verbal systems.

The drastic reduction in the number of nominalizers in the younger informants’ speech, needless to say, provides supporting evidence for principle (a). In fact, we also found cases conforming to principle (b). Specifically, our informant D employs the periphrastic adverbial form *mani* ‘very’ in place of the emphatic reduplication strategy to express degree of intensity.

Such changes are not confined to the realm of derivational morphology. A reduction of inflectional morphology is also found in several of our informants’ speech displaying a pattern of varying degrees of loss, conforming to part of principle (c), shown in (6).

*Contact, Attrition, and Structural Shift in Oroqen*

- (6) a.    tarı           məmi       mənaxa       suxə-rə-n  
           3sg.           self        self:refl.     hit-nonfuture-3sg.  
           ‘He hit himself.’
- b.    bi:            məmi        mənaxa       suxə-rə-n  
           1sg.           self        self:refl.     hit-nonfuture-3sg.  
           ‘I hit myself.’
- c.    ji:            məmi        mənaxa       suxə-rə-n  
           2sg.           self        self:refl.     hit-nonfuture-3sg.  
           ‘You hit yourself.’
- d.    tarı           mUrın-mə       suxə-rə-n  
           3sg.           horse-def.:acc.   hit-nonfuture-3sg.  
           ‘He hit the horse.’
- e.    bi:            mUrın-mə       suxə-m  
           1sg.           horse-def.:acc.   hit-nonfuture:1sg.  
           ‘I hit the horse.’
- f.    ji:            mUrın-mə       suxə-rə-n  
           2sg.           horse-def.:acc.   hit-nonfuture-3sg.  
           ‘You hit the horse.’

This informant is in her late 50s, and the data in (6) were collected in July 1997. She has lost most of her subject-verb agreement, using the third-person singular non-future form for all persons. Notice that she did use the first-person singular verb form in (6e). The examples in (6d-f) were collected after we reminded her of the agreement rule by feeding her the correct forms. Even so, she still used the third-person singular non-future form for the second person as is shown in (6f). However, she does retain the case marking system. She used the accusative markers on the noun ‘horse’ unflinching. In contrast, a slightly younger informant, who is 44 years old, has lost not only subject-verb agreement, but part of the case system as well, which is illustrated in the examples given in (7-10).

- (7) a.    bi:            dʒəktə-yə       dʒək-tə-n/dʒək-tə-m  
           1sg.           food-indef.:acc.   eat-nonfut.-3sg./eat-nonfut.-1sg.  
           ‘I am eating (food).’
- b.    ji:            dʒəktə-yə       dʒək-tə-n  
           2sg.           food-indef.:acc.   eat-nonfuture-3sg.  
           ‘You are eating (food).’
- c.    nənin         dʒəktə-yə       dʒək-tə-n  
           3sg.           food-indef.:acc.   eat-nonfuture-3sg.  
           ‘He is eating (food).’

Fengxiang Li

- d.    bu:           dʒəktə-yə       dʒək-tə-n  
       1pl.:excl. food-indef.:acc. eat-nonfuture-3sg.  
       ‘We are eating (food).’
- e.    biti          dʒəktə-yə       dʒək-tə-n  
       1pl.:incl. food-indef.:acc. eat-nonfuture-3sg.  
       ‘We are eating (food).’
- f.    ʃi:           dʒəktə-yə       dʒək-tə-n  
       2pl.        food-indef.:acc. eat-nonfuture-3sg.  
       ‘You are eating (food).’
- g.    taril         dʒəktə-yə       dʒək-tə-n  
       3pl.        food-indef.:acc. eat-nonfuture-3sg.  
       ‘They are eating (food).’

The examples in (7) show that almost all of the agreement markers between subject and verb are lost in this informant’s speech. The only exception is the first-person singular in (7a), which she uses some of the time. Curiously, the same behavior can be observed in her use of the first-person plural exclusive agreement marking, which is demonstrated in (8).

- (8) a.    bi:           dʒanda-rə-n  
       1sg.        sing-nonfuture-3sg.  
       ‘I am singing.’
- b.    ʃi:           dʒanda-rə-n  
       2sg.        sing-nonfuture-3sg.  
       ‘You are singing.’
- c.    nɔnin        dʒanda-rə-n  
       3sg.        sing-nonfuture-3sg.  
       ‘He is singing.’
- d.    biti          dʒanda-rə-n  
       1pl.:incl. sing-nonfuture-3sg.  
       ‘We are singing.’
- e.    bu:           dʒanda-rə-n / dʒanda-rə-wun  
       1pl.:excl. sing-nonfuture-3sg./sing-nonfuture-1pl.  
       ‘We are singing.’
- f.    ʃu:           dʒanda-rə-n  
       2pl.        sing-nonfuture-3sg.  
       ‘You are singing.’

*Contact, Attrition, and Structural Shift in Oroqen*

- g.     taril           dzanda-rə-n  
        3pl.           sing-nonfuture-3sg.  
        ‘They are singing.’

Clearly, all of the agreement markers are lost in (8a-g), except for (8e) in which case the informant sometimes put the agreement marker in. When asked what the difference is between the two forms with and without the agreement marker, the informant explains that older people prefer forms with the agreement marker, and it is too verbose to talk that way. This indicates that for speakers like her, it is not a matter of imperfectly learning the grammatical system of their L1. They use certain structures by choice, which is counterevidence against some of the claims made in the literature on language attrition situations (e.g. Dorian 1989). In terms of case markings, the informant seems to have a preference for the indefinite accusative case marker to the definite accusative case marker. Notice that in (7), the case marking used on the direct object ‘food’ is the indefinite accusative marker -yə. This is no accident, which can be proven by the examples given in (9).

- (9) a.     nənin           dzaxal                   ʃilki-rə-n  
           3sg.           thing                   wash-nonfuture-3sg.  
           ‘He is washing something.’
- b.     nənin           təti                     ʃiki-rə-n  
           3sg.           clothes                 wash-nonfuture-3sg.  
           He is washing clothes.’
- c.     nənin     min-ŋi     təti-yi                 ʃiki-rə-n  
           3sg.     1sg.-poss. clothes-indef.:acc. wash-nonfuture-3sg.  
           ‘He is washing my clothes.’
- d.     nənin     min-ŋi     təti-wə                 ʃiki-rə-n  
           3sg.     1sg.-poss. clothes-def.:acc. wash-nonfuture-sg.  
           ‘He is washing my clothes.’
- e.     bi:         min-ŋi     təti-yi                 mənəxan  
           1sg.     1sg.-poss. clothes-indef.:acc. 1sg.:refl.  
           ʃiki-m  
           wash-nonfuture:1sg.  
           ‘I am washing my own clothes.’

In (9a) and (9b), there is no accusative marking on the direct object ‘thing’ and ‘clothes’ respectively. When a determiner is added, the informant produced the sentences given in (9c-e). The determiner seemed to have some sort of triggering effect inducing the use of the accusative markers in all of the three examples. However, the speaker apparently has a preference for the indefinite accusative

marker over the definite one, though she switches between the two shown in (9c) and (9d) despite the context requiring the definite accusative marker. Also, in (9e) we again see the first-person singular agreement marker on the verb showing up. It is interesting to note that a deletion of the consonant in the coda position of the initial syllable in the verb *ʃilki-rə-n* ‘wash’ occurred in (9b-e), a phenomenon observed in other informants as discussed earlier in the case of nominalizers. The missing accusative marking in (9a) is not an isolated case. In fact, this informant and some other informants of a comparable age routinely leave out accusative markers, which is shown in the examples in (10).

- (10) a.    *bi:*            *araxi*                *im-nə-n*  
           1sg.          wine                    drink-nonfuture-3sg.  
           ‘I am drinking wine.’
- b.    *ʃi:*            *araxi*                *im-nə-n*  
           2sg.          wine                    drink-nonfuture-3sg.  
           ‘You are drinking wine.’
- c.    *nənin*        *araxi*                *im-nə-n*  
           3sg.          wine                    drink-nonfuture-3sg.  
           ‘He is drinking wine.’
- d.    *biti*         *araxi*                *im-nə-n*  
           1pl.:incl. wine                drink-nonfuture-3sg.  
           ‘We are drinking wine.’
- e.    *bu:*            *araxi*                *im-nə-n*  
           1pl.:excl. wine                drink-nonfuture-3sg.  
           ‘We are drinking wine.’
- f.    *nəntin/taril* *araxi*            *im-nə-n*  
           3pl.          wine                    drink-nonfuture-3sg.  
           ‘They are drinking wine.’

The examples in (10) demonstrate that for this informant, the accusative case marking is no longer obligatory. In fact, she only put it in infrequently. Notice also that in (10) all of the agreement markers have been leveled, reducing them to the default choice of the third-person singular form. It is safe to say that this particular informant has lost most of the case distinctions. When asked about the differences between *mɔː-duki*, *mɔː-laːk*, *mɔː-li*, and *mɔː-ki*, the informant said that they all meant the same thing without hesitation, despite the fact that in each case we have the root morpheme *mɔː* ‘tree’ plus a different case marker: *-duki* ‘ablative (from)’, *-laːk* ‘loco-ablative (from inside)’, *-li* ‘prolative (pass by, along the side)’, and *-ki* ‘allative (to)’. None of these cases is in the grammatical system of this informant. This is also true of another informant comparable in age to this

*Contact, Attrition, and Structural Shift in Oroqen*

informant. Both speakers learned Oroqen as their first language at home. They did not learn Chinese until they started elementary school at the age of seven or eight. It should be pointed out that the language of instruction from grade one through grade five was Oroqen. This means that their dominant language was Oroqen up until they were around 15 years of age, indicating that they lost those forms in a very short period of time, about three decades. A systematic pattern emerges among the informants we have worked with over the years. Those who are 60 or older still have the grammatical system completely intact. All of the older speakers we have worked with so far never failed to use the right agreement markers, and they always readily rejected any violations of the agreement rules. The following paradigm was elicited from a 71-year-old informant in Baiyina in Huma County.

- (11) bi: dzanda-m  
1sg. sing-nonfuture:1sg.  
'I am singing.'
- ʃi: dzanda-ni  
2sg. sing-nonfuture:2sg.  
'You are singing.'
- nɔnin dzanda-rə-n  
3sg. sing-nonfuture-3sg.  
'He is singing.'
- biti dzanda-rə-w  
1pl.:excl. sing-nonfuture-1pl.:excl.  
'We are singing.'
- miti dzanda-ra-p  
1pl.:incl. sing-nonfuture-1pl.:incl.  
'We are singing.'
- su dzanda-ra-y  
2pl. sing-nonfuture-2pl.  
'You are singing.'
- nɔtin dzanda-ra  
3pl. sing-nonfuture:3pl.  
'They are singing.'

The data from speakers with varying degrees of proficiency in Oroqen presented so far show a clear tendency of the younger speakers moving away from inflectional devices and favoring unmarked forms, showing a preference for periphrastic constructions over synthetic ones and for lexemes over bound morphemes to express grammatical relations. The data reveal a restructuring and

reconfiguration of morphological and syntactic structures which display characteristics that provide some insight into longitudinal aspects of language obsolescence. Further investigation is needed to answer some intriguing questions. For instance, we do not fully understand why when subject-verb agreement is lost, second person seems to go first, followed by first person. It is possible that the first person not only has a higher frequency of occurrence but has egocentric prominence as well, resulting in its being more salient, which provides a certain degree of resistance to loss. Why third-person singular is chosen as the default agreement for all persons is unclear. One possible explanation could be that third person is less deictic. Consequently, it is relatively more stable (Lenore Grenoble p.c.). It is also unclear why the definite accusative case marker is lost first. It is likely that the definite accusative case is more cognitively demanding in production since it is semantically more complex. One can argue that the unmarked form is normally retained or lost last. This means that the indefinite accusative case is unmarked with respect to the definite accusative case and the third-person singular agreement marking is unmarked in relation to the other agreement markers. One of our informants once commented that when kids learn Oroqen, they initially only use the third-person singular agreement forms of all verbs. It is possible that frequency of occurrence, saliency, and functional load all have a role to play in the retention and rate of loss of the grammatical forms. It would be of benefit to sort out what goes first and why in terms of the various grammatical components of an obsolescing language. It is equally important to answer the questions of what factors led to those changes and why the changes took place in such a remarkably short period of time. Answers to such questions would help unravel the complexity of language obsolescence taking place in a number of other seriously endangered languages in China.

As Saliger and Vago (1991) pointed out, many, though not all, of the linguistic changes attendant to attrition are simplificatory in nature. This is definitely true in the case of Oroqen. The patterns evident in this simplification process are remarkably similar to those that have already been documented elsewhere (e.g. Schmidt 1985) as is mentioned earlier in this section. However, not all of the changes in Oroqen are explainable through such attrition-induced internal motivations. I argue that, in the case of Oroqen, some of the changes are externally motivated.

### **3. Contact-Induced Changes**

Oroqen has had massive contact for an extended period of time with several genetically related and non-genetically related languages, such as Chinese (Sinitic), Dagur (Mongolic), Ewenki (Machutungusic), and to some extent Russian (Slavic) in the Lesser Hinggan Mountain region. It is imminently clear that for all speakers (except a few elderly speakers), Oroqen is at best a second language, which is being rapidly replaced by Mandarin Chinese. This section is an attempt to identify the changes taking place as speakers replace one language by another, with an emphasis on the role of the contact languages upon the

*Contact, Attrition, and Structural Shift in Oroqen*

phonological, morphological, and syntactic changes which take place, as well as the rate, extent, and order in which these changes occur. The situation Oroqen offers in this respect is quite complex because all the fluent Oroqen speakers are multilingual, at least in Oroqen, Mandarin Chinese, Dagur, and, most importantly, some other Tungusic variety, such as Solon or Evenki.

Among the linguistic effects common to situations of language contact are convergence, loss of morphological and syntactic complexity, transfer, interference, and an overall increase in semantic transparency (Dorian 1989). In the case of Oroqen, we observe borrowing and acts of reception. That is to say, some of the structural changes in this moribund language may be the result of influence from linguistic aspects of the dominant languages. Before the encroachment of Mandarin Chinese in the 1950s, Oroqen had had lengthy and persistent contact with the Dagurs resulting in the borrowing of certain grammatical structures. A case in point is the emphatic reduplication strategy to mark intensity, which is no longer present in the grammatical system of younger speakers of Oroqen as is discussed in section two above. Whaley and Li (2000) demonstrated convincingly that Oroqen borrowed the emphatic reduplication strategy from Dagur, a Mongolic language.

Dagur has a formally identical reduplication strategy which copies the first syllable and inserts [b] or [m] in the coda position of the prefix (Zhong 1982), as is illustrated in the examples in (12).

|      |         |         |              |                  |
|------|---------|---------|--------------|------------------|
| (12) | xula:n  | ‘red’   | xub xula:n   | ‘thoroughly red’ |
|      | tʃiya:n | ‘white’ | tʃim tʃiya:n | ‘very white’     |
|      | dasuŋ   | ‘sweet’ | dab dasuŋ    | ‘really sweet’   |
|      | səru:ŋ  | ‘cool’  | səb səru:ŋ   | ‘really cool’    |
|      | xorduŋ  | ‘fast’  | xob xorduŋ   | ‘very fast’      |

Like Oroqen, Dagur employs reduplication to indicate intensity. However, the process in Dagur is fully productive and operates on adjectives denoting different sorts of properties, not just colors.

Dagur speakers have been in contact with Oroqen speakers for centuries, in a symbiotic relationship. The traditionally nomadic Oroqen relied on the sedentary Dagur for certain agricultural goods, while in exchange supplying the Dagur with pelts and meat. The commercial relationship, while mutually beneficial, established Dagur as the dominant language, and it became the norm for Oroqen speakers to learn to speak Dagur.

Consequently, the Oroqen lexicon has taken on many Dagur words, and Oroqen grammar has borrowed from Dagur, particularly in the area of derivational morphology. Thus, the borrowing of reduplication can be seen as part of a more general Dagur influence on Oroqen grammatical structure.

The sort of structural influence that Dagur has had on Oroqen requires an extended period of relatively intimate contact. It is useful here to review the history of these groups which points to just this type of interaction.

Both the Dagurs and the Oroqens are believed to originate from the region north of the Amur River in present-day Russia. They both crossed the mighty Amur River several hundred years ago (most likely in the 1600s) and spread over the Greater and Lesser Hinggan Mountains in Manchuria. Janhunen (1997) suggests that migrations of small Dagur populations occurred in tandem with the Oroqen and Solon migrations. Regardless, it is widely accepted that all these groups have co-existed harmoniously in Inner Mongolia and Northeast of China for several centuries. Since the borrowing being discussed here is structural rather than lexical in nature, massive bilingualism on the part of the Oroqen speakers in the lending language persisting over a long period of time is a crucial prerequisite.

Another social factor frequently invoked in accounting for borrowings across languages, especially languages belonging to different families, is the prestigious status of the source language. As Moravcsik (1978) puts it, in perhaps an overstatement, “nothing can be borrowed from a language which is not regarded [as] prestigious by speakers of the borrowing language.” The Dagur people have historically held the necessary position of prestige: according to Janhunen (1997), “since Qing times (1644-1911), the Dagurs have been known as an ethnic group interested in acquiring higher learning through dominant languages such as Manchu and Chinese.” This may help explain why the Qing emperor entrusted the Dagurs with the control of the diaspora army sent to guard Chinese borders against the potential invaders in bordering areas.

Dagur is not the only contact language that has had an impact on the structure of Oroqen. In more recent times, Mandarin Chinese has undoubtedly influenced Oroqen. Although the period of time that Mandarin Chinese has been in intense contact with Oroqen is relatively short, it seems to have had a strong impact on its grammatical structure. We found that for most speakers of Oroqen, the plural marker is no longer required, which could be the consequence of Chinese influence (cf. Grenoble and Whaley to appear). Although Hu (1986) documented both *-l* and *-sal* as plural markers in Oroqen, they are no longer readily attested in any of the Oroqen dialects. As Grenoble and Whaley (to appear) pointed out, a zero morpheme is the unmarked, preferred plural marking in all of the Oroqen dialects. They also noted that Oroqen is the only Northwestern Tungusic language in which unmarked plurals are more common than suffixation. Assuming that Hu’s (1986) description is accurate, it is noteworthy that Oroqen has moved further on its path to losing the plural marker. We have yet to elicit any naturally occurring examples of Oroqen in which the plural markers are employed. The informants we have worked with from southeastern, western, and central Oroqen dialect regions occasionally very reluctantly accepted the plural marker *-l* and *-sal* in a highly restricted number of lexical items, most of which denote animate beings with a high frequency of occurrence. The only informant who readily accepted forms with the plural markers *-l* and *-sal* was from the northeastern Oroqen

dialect region, specifically from Baiyina. Some of the examples are: *kumaxa-l* ‘deer’, *utə-l* ‘sons’, *ilga-l* ‘flowers’, *bəyə-səl* ‘persons’, *ahi-səl* ‘women’, *utə-səl* ‘sons’. Even for her, the preferred form is the analytic construction exemplified by *baran kumaxa* ‘many deer’. It is highly likely that this loss is contact induced.

#### **4. Conclusion**

These phenomena pose the questions of what factors led to these changes and why the changes took place in such a remarkably short period of time. The complex set of conditions responsible for the restructuring are only partly explainable by internally and externally motivated principles proposed in the literature. A profound understanding of the situation must take account of the dynamic changes that take place in not only linguistic structures but in social conditions as well.

The conditions responsible for the changes in Oroqen are mostly sociological in nature. Oroqen was surrounded by several languages for centuries without suffering any attrition, which is attributable to the cohesion of the community, its self-imposed rule forbidding intermarriages with other ethnic groups and its strong adherence to traditional lifestyle and values. However, dramatic social changes in the 50s and 60s quickly resulted in Oroqen’s alarmingly rapid ceding to the superstratum language of Mandarin Chinese. The construction of a railroad, massive Chinese migration to the area due to the development of the logging and mining industries, resettlement of the Oroqens, schooling in Chinese, extensive intermarriages with nearby ethnic groups, abandonment of traditional values and lifestyle, and the local government’s lack of interest in maintaining the Oroqen language and cultural traditions led to a situation characterized by an increasing use of Mandarin Chinese in more and more situations, which eventually brought about a proficiency continuum determined by age among those who still speak Oroqen. The diminishing use of Oroqen severely limits the younger generation’s exposure to the language, which is the major cause for their imperfect learning of their obsolescing first language. It is quite obvious that the Oroqen language is on an irreversible course to its demise, and we are far short of attaining a full picture of the confluence of factors that led to its current sad state of affairs.

In this paper I only presented a few salient characteristic features. I would like to emphasize that what is discussed here is barely a sampling which should not be construed as an attempt to provide a full picture, but rather as a requisite component to attain the ultimate goal of uncovering all the attrition and contact phenomena in this language undergoing rapid attrition, contributing to the general effort to achieve an adequate account of changes attendant upon the attrition process.

#### **References**

Doerfer, G. 1983. Das Birare. *Journal de la Société Finno-Ougrienne* 78:7-25.

- Dorian, Nancy C., ed. 1989. *Investigating Obsolescence: Studies in Language Contraction and Death*. Cambridge: Cambridge University Press.
- Grenoble, Lenore A., and Lindsay J. Whaley, eds. 1998. *Endangered Languages: Current Issues and Future Prospects*. Cambridge: Cambridge University Press.
- Grenoble, Lenore A., and Lindsay J. Whaley, eds. To appear. The Case for Dialect Continua in Tungusic: Plural Morphology. In D. A. Holisky and K. Tuite, eds., *The Non-Slavic Languages of the Former USSR: Linguistics Studies: In Honor of Howard Aronson*. Chicago: Chicago Linguistics Society.
- Han, Y., and S. Meng. 1993. *Elunchun Yu Han Yu Duizhao Duben* [An Oroqen-Chinese Reader]. Beijing: Zhongyang Minzu Xueyuan Chubanshe [Central Nationalities Institute Press].
- Hill, Jane H. 1980. Language Death in Uto-Aztecan. Paper for the Conference on Uto-Aztecan Linguistics. Albuquerque. Ms.
- Hu, Zengyi. 1986. *Elunchun Yu Jianzhi* [A Grammatical Sketch of Oroqen]. Beijing: Minzu Chubanshe [Nationalities Press].
- Janhunen, Juha. 1996. *Manchuria: An Ethnic History*. Mémoires de la Société Finno-Ougrienne 222. Helsinki.
- Janhunen, Juha. 1997. The Languages of Manchuria in Today's China. *Northern Minority Languages: Problems of Survival*. Senri Ethnographical Studies 44:123-146.
- Janhunen, Juha, Jingxue Xu, and Yucheng Hou. 1989. The Orochen in Xinsheng. *Journal de la Société Finno-Ougrienne* 82:145-169.
- Li, Bing. 1992. Elunchunyu de Yuanyin Hexie [Oroqen Vowel Harmony]. *Minzu yuwen* 6:47-53.
- Li, Bing. 1996. *Tungusic Vowel Harmony: Description and Analysis*. Amsterdam: Universiteit van Amsterdam.
- Li, Fengxiang. 1996. A Report on the Study of the Consonant Systems of Two Manchu-Tungusic Languages in China: Evenki and Oroqen. Paper presented at the Annual Meeting of the Linguistic Society of America, San Diego.
- Li, Fengxiang, and Lindsay J. Whaley. 1998. The Rise and Fall of Derivational Morphology: The Case of Reduplication in Oroqen. *Proceedings of the Thirty-Fourth Regional Meeting of the Chicago Linguistic Society*.
- Li, Fengxiang, and Lindsay J. Whaley. 2000. Emphatic Reduplication in Oroqen and Its Altaic Context. *Linguistics* 38(2):355-372.
- Li, Shulan. 1981. Elunchun-yu cihui shulue [A brief description of the vocabulary of Oroqen]. *Minzu yuwen* 5:405-26.
- Maher, Julianne. 1991. A Crosslinguistic Study of Language Contact and Language Attrition. In H. W. Seliger and R. M. Vago, eds., *First Language Attrition*: 67-84. Cambridge: Cambridge University Press.
- Meng, S. 1993. *Elunchun Min Jian Wenxue* [Oroqen Folk Literature]. Harbin: Heilongjiang Sheng Minzu Yanjiusuo [Nationalities Research Institute of Heilongjiang Province].
- Moravcsik, Edith A. 1978. Language Contact. In J. H. Greenberg, ed., *Universals of Human Language* 1:93-122. Stanford: Stanford University Press.

*Contact, Attrition, and Structural Shift in Oroqen*

- Saxirong. 1981. *Jianming Hanyu Elunchunyu duizhao duben*. [A concise Chinese-Oroqen reader]. Beijing Academy of Social Sciences
- Schmidt, Annette. 1985. *Young People's Dyrbal*. Cambridge: Cambridge University Press.
- Seliger, Herbert W., and Robert M. Vago, eds. 1991. *First Language Attrition*. Cambridge: Cambridge University Press.
- Shirokogorov, S. M. 1923. *Anthropology of Northern China*. Shanghai.
- Shirokogorov, S. M. 1929-1933. *Social Organization of the Northern Tungus*. Shanghai: The Commercial Press Limited.
- Shirokogorov, S. M. 1935. *Psychomental Complex of the Tungus*. London.
- Shirokogorov, S. M. 1944. *A Tungus Dictionary: Tungus-Russian and Russian-Tungus*. Tokyo: Minzo kugaku Kyokai.
- Thomason, Sarah Grey and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Weinreich, Uriel. 1953. *Languages in Contact*. The Hague: Mouton.
- Whaley, L. J., L. A. Grenoble, and F. Li. 1999. Revisiting Tungusic Classification from the Bottom Up: A Comparison of Evenki and Oroqen. *Language* 75.
- Whaley, Lindsay J., and Fengxiang Li. 1998. The Suffix *-kan* in Oroqen. *Studies in Language* 22:181-205.
- Whaley, Lindsay J., and Fengxiang Li. 2000. Oroqen Dialects. *Central Asiatic Journal* 44(1):105-130.
- Zhang, Xi. 1996. Vowel Systems of the Manchu-Tungus Languages in China. Unpublished dissertation, University of Toronto.
- Zhang, Y., B. Li, and X. Zhang. 1989. *The Oroqen Language*. Changchun: Jilin University Press.
- Zhong, Suchun. 1982. *Dagur Yu Jianzhi* [A Grammatical Sketch of Dagur]. Beijing: Minzu Chubanshe [Nationalities Press].

Fengxiang Li  
English Department  
California State University, Chico  
Chico, CA 95929-0830

fli@csuchico.edu



## **Cross-Language Perception of Syllable Affiliation: Effects of Voicing and Language Background\***

BYUNG-JIN LIM<sup>1</sup>, KEN DE JONG<sup>2</sup>, and KYOKO NAGAO<sup>2</sup>  
*The University of Wisconsin-Madison<sup>1</sup> and Indiana University<sup>2</sup>*

### **0. Introduction**

This paper examines the effect of linguistic experience in listeners' perception of syllabic affiliation. While there is a large amount that we know about the second language acquisition of paradigmatic contrasts, there is relatively little parallel literature concerning the acquisition of syntagmatic contrasts. To do this, we compared first language and second language perceptions of rate-varied stimuli, which have been noted in previous studies to vary in perceived syllabic structure. We also in parallel compared first and second language perception of 'voiced' and 'voiceless' stops in the same corpus, in order to ascertain how well the listeners would do with a phonemic contrast in the same type of data. In general, this study finds that second language learners are very much like native listeners in the perception of syllabic affiliation, but not in the perception of voicing.

Stetson (1951, also partially documented in Tuller and Kelso 1991) noted that singleton coda consonants (VC) appear to modulate into onset consonants (CV) when repeated at relatively fast rates. Thus, with CV forms, rate increases seem to generate a continuum ranging from clearly perceived codas to clearly perceived onsets. Hence the current study examines perceptual responses to such repetitive productions to ascertain the degree to which speakers of languages with different syllabic inventories actually do perceive such resyllabification.

In addition, in order to examine the generalizability of the resyllabification phenomenon across segments with different temporal characteristics, the production corpus varied in the 'voicing' of the resyllabified consonants. Hence, we are able to examine the same listeners' responses to voicing contrasts in the same data. Voice onset time (VOT) is a well-known acoustic attribute distinguishing the voicing contrast in initial stops (Lisker and Abramson 1964). Changes in speech rate affect the range of VOT values, particularly for voiceless stops (Miller et al. 1986; Miller and Volaitis 1989; Volaitis and Miller 1992; Pind 1995; Kessinger

---

\* Research supported by the NIDCD (grant R03 DC04095-01A2) and by the NSF (grant BCS-9910701).

and Blumstein 1997, 1998). Either due to some relational property of the speech itself, or due to listeners' learning of these rate effects, the perceptual boundary between voiced and voiceless stop consonants accordingly shifts toward shorter VOT values as syllable duration decreases (Summerfield 1981, Miller et al. 1986, Miller and Volaitis 1989, Volaitis and Miller 1992).

Non-native listeners were speakers of languages for which either voicing or syllabification is grammatically restricted. One group of listeners was native speakers of Japanese, a language which does not allow post-vocalic obstruents without a following vowel. Japanese does allow geminate consonants which are generally treated as acting as the coda of the preceding syllable, and among younger speakers, these geminates can contrast in voicing. Hence, Japanese seems to be a case of a language for which an obstruent appearing as a singly affiliated coda is not native, but in which, nevertheless, voicing contrasts in post-vocalic position are possible. The second group of listeners was native speakers of Korean. Korean, in contrast to Japanese, does allow words ending in an obstruent. However, also in contrast with Japanese, consonants which are co-syllabic with the preceding vowel do not contrast in voicing, due to a phonological neutralization rule. A third group of native English listeners acted as a control group. All three languages have a voicing contrast in onset position, though the exact contrast differs across all three languages. Table 1 summarizes these facts.

**Table 1.** Cross-language comparison: syllable affiliation and voicing contrast

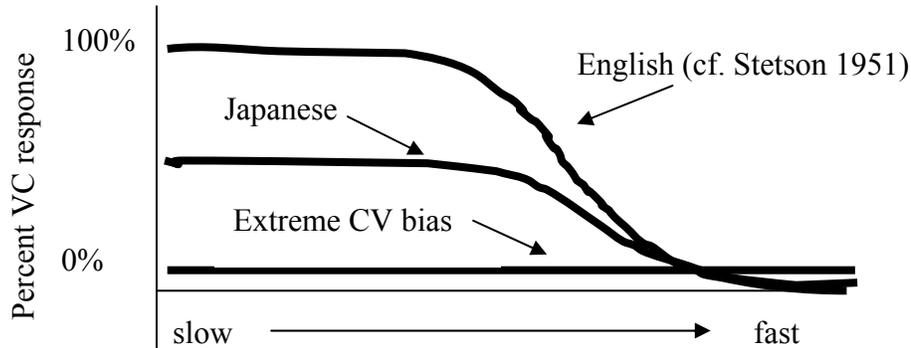
|       | English |           | Japanese      |           | Korean      |           |
|-------|---------|-----------|---------------|-----------|-------------|-----------|
|       | voiced  | voiceless | voiced        | voiceless | voiced      | voiceless |
| Onset | b       | p         | b             | p         | b           | p         |
| Coda  | b       | p         | geminate only |           | neutralized |           |

## 1. Hypotheses

### 1.1. Syllabification Perception

We hypothesize that listeners' perception of syllable affiliation depends on the syllabic inventory of their native language. Specifically, since Japanese does not allow syllable final stops in coda positions while English and Korean do, Japanese listeners will exhibit a bias toward perceiving syllable structures as CV. In the extreme case, Japanese listeners might not hear VC forms at all. Another possibility would be that Japanese listeners are unsure about how to categorize English VC forms and have identification functions which start around 50% at slow rates, and shift as stimuli sound more like canonical CV's. Coupled with Stetson's observations, then, we might expect the pattern of results in Figure 1.

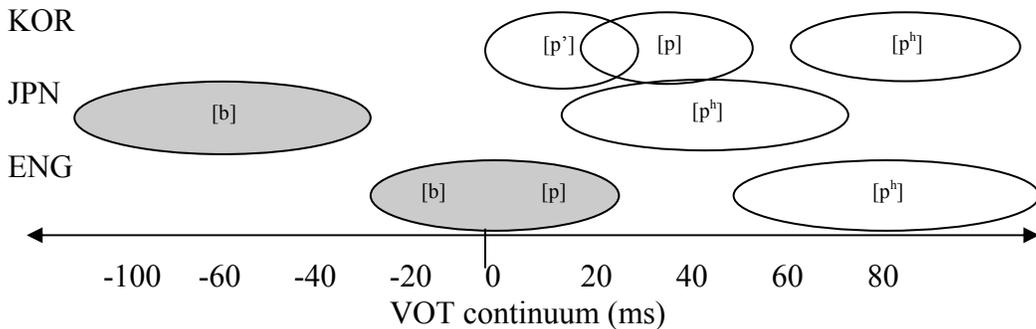
**Figure 1.** Prediction for perception of syllable affiliation



**1.2. Voicing Perception**

We hypothesize that voicing perception will be influenced by the perceivers’ native voicing categories. For onsets, Korean has a three-way distinction in voiceless stops.<sup>1</sup> Here, the boundary between aspirated and unaspirated stops is higher in VOT than that in English. Japanese has a two-way contrast whose boundary seems to be lower in VOT value than that in English. The categories in VOT value are illustrated in Figure 2. Given the direction of the boundary differences from English, we expect a bias in Japanese listeners against hearing items as voiced, and a bias in Koreans against hearing items as aspirated.

**Figure 2.** Schematic phonetic categorization for voicing on the VOT continuum (Japanese and Korean VOT values from Shimizu 1996)

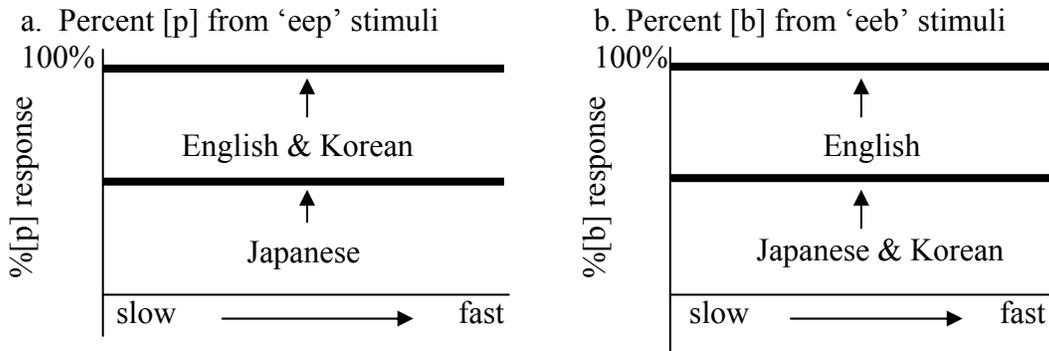


Predictions with respect to the voicing of coda consonants are more complicated. Since Japanese does not have singleton coda consonants, we might expect that Japanese natives would have difficulty perceiving both voiced and voiceless stops, and be simply guessing as to the voicing of coda stops. Also, since Korean exhibits the neutralization of coda voicing toward voicelessness, we might expect a bias toward voiceless labeling. These combined predictions can be schematized as in Figure 3. However, we should also note that younger Japanese speakers are

<sup>1</sup> Intervocally, however, lax stops can be voiced in Korean.

familiar with nativized foreign words which contrast in the voicing of geminate stops, and Korean does contrast voiced (lenis) and voiceless (fortis or aspirated) stops in intervocalic position. Hence, we might find native-like perception of voicing.

**Figure 3.** Predictions for voicing perception



## 2. The Perception Experiment

### 2.1. Methods

#### 2.1.1. Subjects

The subjects were 21 native speakers of American English, 17 native speakers of Japanese, and 13 native speakers of Korean. They were all of university age with normal hearing and no linguistic training. All were recruited through responses to posted advertisements and were paid \$10 each for their participation.

#### 2.1.2. Stimuli

Stimuli were extracted from a repetitive production experiment in which four English talkers produced voiced and voiceless labial stop onsets and codas at gradual increasing tempi from 450 to 150 ms/syllable controlled by metronomes (for details concerning the production of the forms, see de Jong 2001). Each stimulus included three repetitions of one of each of the four monosyllables, 'pea', 'bee', 'eep', and 'eeb'. 404 stimuli were presented to each listener group.

#### 2.1.3. Procedures

After listening to recordings through headphones, subjects were asked to identify what they heard by clicking on a computer monitor. There were four buttons, labeled 'pea', 'bee', 'eep', and 'eeb'. Prior to experimentation, a practice run of 8 trials was given. During experimentation, a five-minute break was given after every 100 trials.

## 2.2. Results

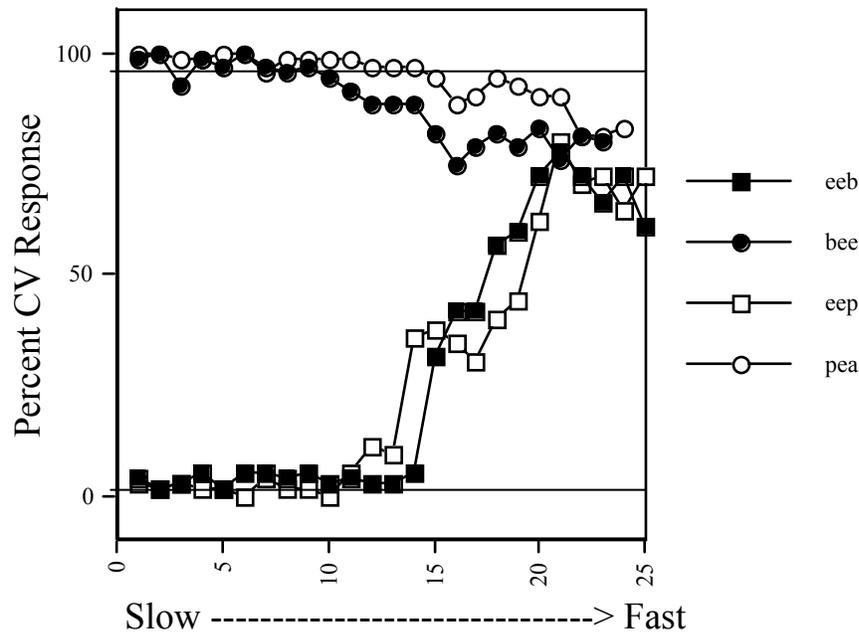
### 2.2.1. Perception of Syllable Affiliation

To replicate Stetson's observations, we first turn to the English subjects. Figure 4 shows the results of perception of syllable affiliation for CV and VC structures by

### Cross-Language Perception of Syllable Affiliation

English listeners. As shown in the top of Figure 4, there is little effect of speech rate on the perception of CV syllable structure. Listeners identified CV productions as CV forms. We do note a slight decrease in CV responses for CV stimuli, though the amount of change is quite small (less than 20%). In the perception of VC tokens, as shown in the bottom of Figure 4, we note a clear shift in identification from VC to CV as speech rate increases. This pattern agrees with Stetson (1951). Voicing differences do not seem to affect the pattern of identification. Both ‘eeb’ and ‘eep’ stimuli elicit ogival functions, though there is an earlier and more gradual shift in ‘eep’ stimuli.

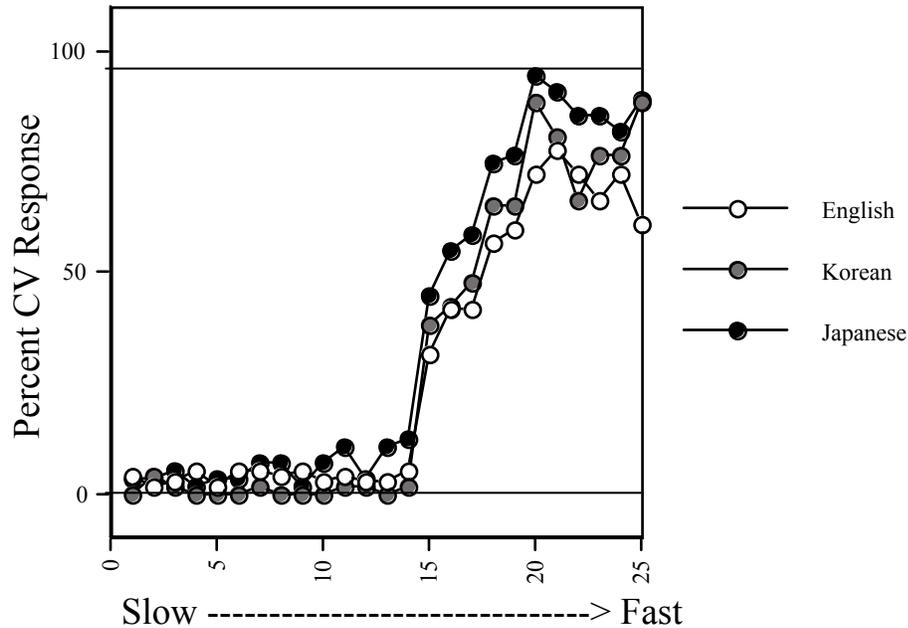
**Figure 4.** Identification functions for CV (circles) and VC (squares) stimuli. The horizontal scale indicates the location of the stimulus in the trial from the slow rate beginning (left) to the fast rate ending (right).



#### 2.2.2. Cross-Language Perception of Syllable Affiliation

Figure 5 shows the identification of syllable affiliation for non-native listeners. Responses plotted here are only for ‘eeb’ stimuli. All three groups exhibit resyllabification, and the category shift seems to occur at nearly the same location in the continuum. Thus, all three response functions are very similar in shape and location. There is a difference in the three groups’ responses to faster rates, however. Especially Japanese listeners were more likely to label fast rate tokens as CV’s. Hence, there is an apparent bias toward CV labeling, but only at fast repetition rates. Responses to ‘eep’ tokens were similar.

**Figure 5.** Identification functions for VC tokens for native speakers of English, Japanese and Korean



### 2.2.3. Cross-Language Perception of Voicing Contrast

For comparison, we turn to the perception of the voicing contrast. For reasons noted above, we consider voicing perception for stops produced as onset stops separately from stops produced as codas. The results of voicing perception for stops produced as CV's are shown in Figure 6.<sup>2</sup> Examining the English listeners first (unfilled symbols), one notes that native listeners are generally very good at identifying the voiceless consonants, while voiced consonants at fast rates tend to be misidentified as voiceless. These general patterns are visible in the non-native listeners as well, though with some notable differences. Korean listeners (gray symbols) tended to hear some voiceless tokens as voiced, and Japanese listeners (black symbols) tended to hear voiced tokens as voiceless. These effects are true even of slow rate stimuli. These are what one would expect as biases induced by the native categories as illustrated in Figure 2. Koreans tend to require more aspiration for voiceless stops, and Japanese tend to require more voicing for voiced stops.

<sup>2</sup> Note the results shown here only include three of the four speakers, since a fourth speaker elicited very different responses.

**Figure 6.** Proportion of voiceless responses to phonemically voiced and voiceless stimuli by the three language groups

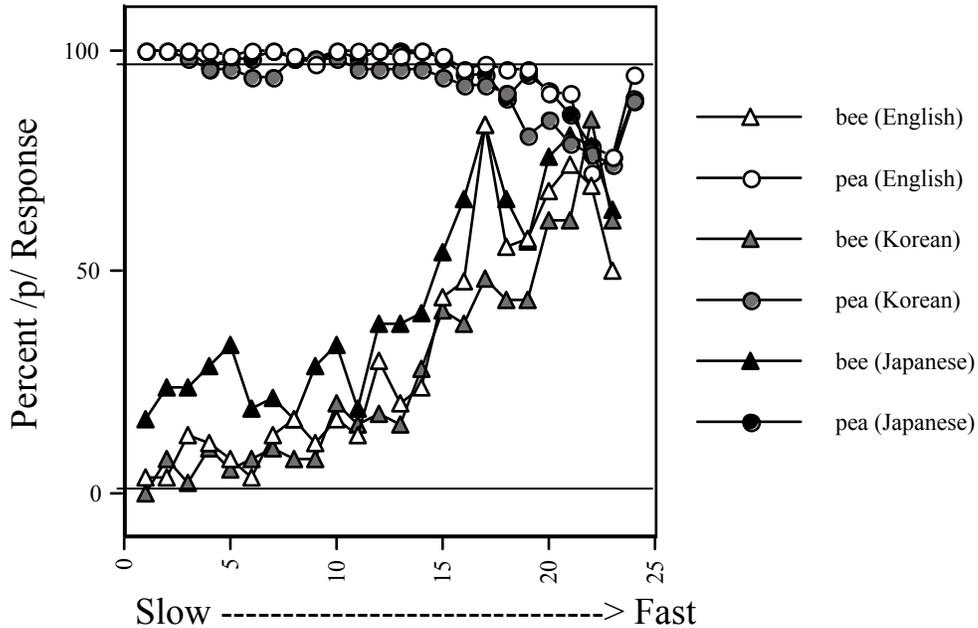
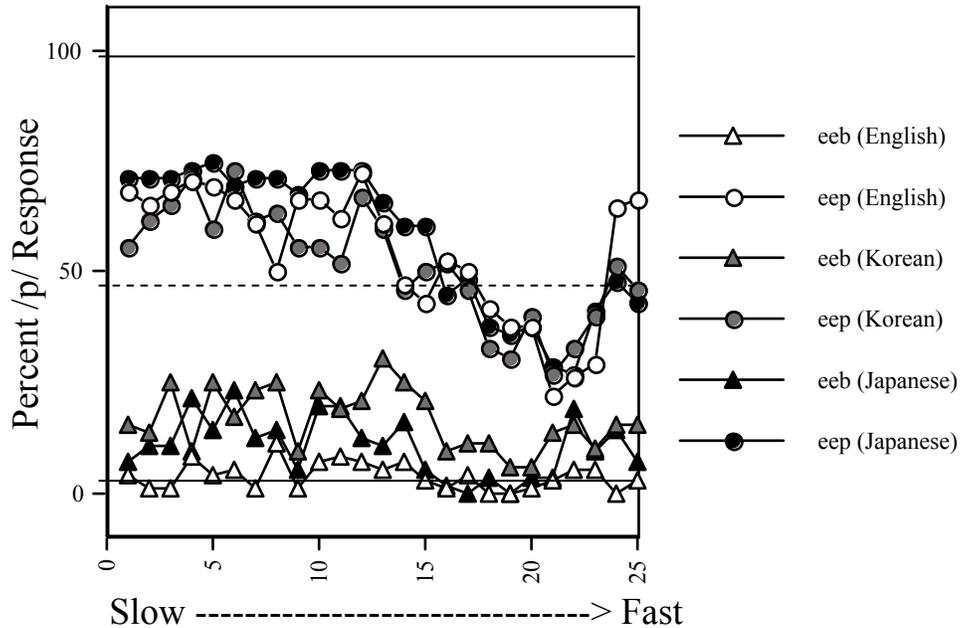


Figure 7 plots the identification of voicing for stops produced as codas. Here, English listeners were generally good at identifying voiced stops, but were very poor at identifying voiceless stops. Fast rates tended to induce even more mis-identifications of voiceless as voiced stops. Non-native listeners showed the same asymmetry, and were, in general, very similar to native listeners. However, as with onset stops, we also find differences. While native listeners were very good at identifying voiced coda stops, both non-native groups were less consistent, especially Korean listeners. Korean listeners exhibited identification functions shifted toward chance with respect to native listeners.

**Figure 7.** Proportion of voiceless responses to phonemically voiced and voiceless stimuli by the three language groups.



### 3. Discussion and Conclusions

#### 3.1. Perception of Resyllabification

Regardless of the speakers' native language syllabic inventory, all listener groups showed very similar patterns of perceptual resyllabification as a function of speech rate change. Currently, it is unclear whether this similarity is due more to categorical changes in the articulation of the stimuli, or due entirely to similarities in the perceptual boundaries for the listeners. However, it is remarkable not only that English listeners bear out Stetson's observations, but also that acquirers whose native languages do not have the same inventory of structures exhibit essentially the same perceptual patterns. Thus, it seems that the perception of syllable affiliation seems to be, if not a cross-language universal skill, at least an aspect of production which is very accessible to acquirers.

Beyond this, it is not the case that non-native perceivers are identical to native perceivers. Particularly interesting differences begin to arise in the perception of fast rate utterances. There is a tendency for the non-native perceivers to actually be more likely to label VC's as being resyllabified as CV's. This very subtle bias away from the non-native VC structure seems to fit nicely with a model in which their perception of forms in a non-native language retains a bias toward forms which have close analogies in their native languages. In this case, Japanese-speaking listeners would be more likely to perceive VC forms as resyllabified into

CV forms because CV forms occur in Japanese native vocabulary. However, such an explanation does not fit well with the Korean listeners, for whom VC forms are native. The similarity of the Japanese and Korean listeners with regard to resyllabification suggests a somewhat different explanation. Rather than non-native listeners exhibiting a bias toward native forms, the English listeners are exhibiting resistance to labeling the VC forms as something other than what the speaker intended. That is, English listeners may simply be better at detecting the traces of the fast rate stimuli's VC origin than the non-native listeners. Previous acoustic analyses of the stimuli that the current study used show differences between resyllabified codas and onsets both in the temporal characteristics of the closure and in the spectral quality of the vowel (de Jong 2001). Apparently, native listeners are sensitive to these differences.

### **3.2. Effects of Phonemic Contrasts on Voicing Perception**

As previously mentioned, both Japanese and Korean have a contrast in onset voicing roughly analogous to the English contrast. Thus, it is not surprising that both non-native listener groups are generally good at identifying voicing in the current data. Further, the voicing categories in the three languages are different in detail. For instance, there are typically different boundaries on a VOT continuum for each language. This difference in voicing categorization seems to be reflected in subtle biases in the non-native responses; English /b/ tended to be heard as [p] by Japanese listeners while English /p/ tended to be heard as [b] by Korean listeners.

At first blush, this pattern of results seems to match that for syllabification. Non-native speakers of English generally match native speakers with a small bias toward responding in terms of a native category. However, there is a difference between the two sets of results with respect to rate. Specifically, syllabification biases showed up at fast rates, where there is some uncertainty in the native speaker responses. Voicing biases are detectable at all rates, particularly the Japanese bias toward voiceless responses. Hence, it seems that the existence of the non-native category is detectable even in cases in which native listeners are consistent in their responses.

Considering the voicing contrast in codas, which we expect to be non-native, we find yet another pattern. Non-native listeners are simply less consistent at identifying the voicing of coda stops. This is particularly evident in the Korean listeners, whose identification functions were both shifted toward chance from the native identification functions. This, again, makes intuitive sense if we consider that not having a native contrast will result in less developed identification skills. It is surprising that the Korean listeners had problems with voiceless stops, even though they are said to produce the voiceless variants. We suspect this indicates that the neutralization of voicing in coda position creates in perception a lack of attention to voicing cues such as vowel duration. If this is the best explanation for the current results, it underscores the point made in the previous section concerning syllabification. Despite not having experience with determining the syllabic

affiliation of stops between vowels, both non-native listener groups performed very much like natives of English. Syllabification of obstruents does seem to be remarkably accessible to non-native listeners.

There are a number of additional issues which the current results raise, most particularly concerning the general pattern of voicing identification in onsets and codas. For example, voiceless onsets and voiced codas both are well identified, while voiced onsets and voiceless codas are relatively poorly identified. In both of the latter cases, in addition, the poor identification gets worse as speech rate increases. What is particularly remarkable about this pattern is how similar it is for native and non-native listeners. This leads us to believe that the explanation for these overall patterns of responses will be found in how voicing is expressed in the repetitive speech, and not in peculiarities of the listeners' perceptual categories. We are currently in the process of sorting out the relationship between the numerous cues to voicing and the identification patterns of the three groups.

## References

- de Jong, K. J. 2001. Rate-Induced Resyllabification Revisited. *Language and Speech* 44(2):197-216.
- Kessinger, R. H., and S. E. Blumstein. 1997. Effects of Speaking Rate on Voice-Onset Time in Thai, French, and English. *Journal of Phonetics* 25:143-68.
- Kessinger, R. H., and S. E. Blumstein. 1998. Effects of Speaking Rate on Voice-Onset and Vowel Production: Some Implications for Perception Studies. *Journal of Phonetics* 26:117-28.
- Lisker, L., and A. S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20:384-422.
- Miller, J. L., K. P. Green, and A. Reeves. 1986. Speaking Rate and Segments: A Look at the Relation Between Speech Production and Speech Perception for the Voicing Contrast. *Phonetica* 43:104-15.
- Miller, J. L., and L. E. Volaitis. 1989. Effects of Speaking Rate on the Perceptual Structure of a Phonetic Category. *Perception and Psychophysics* 46:505-12.
- Pind, J. 1995. Speaking Rate, Voice-Onset Time, and Quantity: The Search for Higher Order Invariants for Two Icelandic Speech Cues. *Perception and Psychophysics* 57:291-304.
- Shimizu, K. 1996. *A Cross-Language Study of Voicing Contrasts of Stop Consonants in Six Asian Languages*. Tokyo: Seibido.
- Stetson, R. H. 1951. *Motor Phonetics: A Study of Speech Movements in Action*. Amsterdam: Noth-Holland.
- Summerfield, A. Q. 1981. Articulatory Rate and Perceptual Constancy in Phonetic Perception. *Journal of Experimental Psychology: Human Perception and Performance* 7:1074-95.
- Tuller, B., and J. A. S. Kelso. 1991. The Production and Perception of Syllable Structure. *Journal of Speech and Hearing Science* 34:501-8.

*Cross-Language Perception of Syllable Affiliation*

Volaitis, L. E., and J. L. Miller. 1992. Phonetic Prototypes: Influence of Place of Articulation and Speaking Rate on the Internal Structure of Voicing Categories. *Journal of the Acoustical Society of America* 92:723-735.

Byung-Jin Lim  
Van Hise Hall 1214  
East Asian Languages and Literature Department  
The University of Wisconsin-Madison  
1220 Linden Drive  
Madison, WI 53706

blim3@wisc.edu

Ken de Jong  
Memorial Hall 322  
1021 East 3rd Street  
Linguistics Department  
Indiana University  
Bloomington, IN 47405

kdejong@indiana.edu

Kyoko Nagao  
Memorial Hall 322  
1021 East 3rd Street  
Linguistics Department  
Indiana University  
Bloomington, IN 47405

knagao@indiana.edu



# Speech Reporting Practices and Subjectivities: Evidence from an Amazonian Society

LEV MICHAEL

*University of Texas at Austin*

## 1. Introduction

This paper<sup>1</sup> examines the speech reporting practices (SRPs) of the Nanti, an Arawakan people of the Peruvian Amazon.<sup>2</sup> The goals of this paper are to provide a description of Nanti SRPs from a communicative-functional perspective, and to account for the observed organization of Nanti SRPs from this standpoint.

In particular, I describe the exclusive use of direct speech reporting (DSR) for the reporting of speech, and the use of DSR to represent illocutionary forces, evaluative judgments, subjective orientations, and agency in Nanti discourse. On the basis of this evidence, I argue that Nanti SRPs form part of a practice of representation of subjectivity and agency in discourse that makes use of a non-dualistic model of the person in which agency and subjectivity are represented through reported speech. I argue that the organization of Nanti SRPs can be understood as stemming from the place of SRPs in this system of discursive representation of the person.

I will compare my findings for Nanti SRPs with those of other scholars working in other societies in Latin America and beyond, and note an apparent correlation between SRPs and cultural models or ideologies of the speaking subject.

---

<sup>1</sup> This paper results from fieldwork I have carried out with Chris Beier, and many of the above ideas are derived from discussions with her. This paper has benefited tremendously from discussions with Joel Sherzer and Elizabeth Keating. *Yogari matsigenxa montetonixu yogotagajigax-enara irinijane. Noxanti inti xameti matsigenxa montetonixu.* The shortcomings that remain in this work are my own.

<sup>2</sup> The Nanti live at the headwaters of the Camisea and Timpia Rivers, located in the lowland rainforest of southeastern Peru. There are roughly 600 speakers of Nanti, living in approximately ten shifting settlements. During four fieldwork trips from 1997 to 2000, I spent ten months in the two largest of these settlements, Montetoni and Maranxajari, which have a combined population of approximately 250 individuals.

Nanti is closely related to several other better-known Arawakan languages spoken in southeastern Peru, including: Asháninka, which is estimated to have 15,000-18,000 speakers; Ashéninka, which is estimated to have 18,000-25,000 speakers; and Matsigenka, which is estimated to have 7,000-12,000 speakers (Aikhenvald 1999).

Finally, I will relate my findings to recent work that links speaking practices to forms of subjectivity, and explore the implications of the comparative evidence from Nanti society for research on the relationship between SRPs and forms of subjectivity.

## 2. Nanti Speech Reporting Practices

I will begin by describing the two major characteristics of Nanti SRPs that are relevant to the issues raised: (i) Nanti speakers' exclusive use of direct speech reporting (DSR) to report speech, and (ii) Nanti speakers' use of DSR in settings in which speakers of European languages would typically employ performative verbs of speaking or mental activity verbs to introduce indirectly reported speech or represented thought.

Of these, the exclusive use of DSR for reporting speech might be the most immediately striking aspect of Nanti speaking practices for a speaker of a European language such as English or Spanish.

Imagine the following scene on a path between a Nanti village and the river: a woman on her way to the river passes a child. As she passes, she says to him *Nojate ojaxu* 'I am going to the river'. Minutes later, her husband passes by the same child, looking for his wife, and asks the child *Tyaxa ojataxe?* 'Where did she go?'. The child responds *Oxanti nojate ojaxu* 'She said, "I am going to the river"'. Under these circumstances the child would never say *Oxanti ojate ojaxu* 'She said she is going to the river'. That utterance would be interpreted by a Nanti speaker as 'She (the woman who passed by) said that she (another woman) was going to the river'.

Abandoning hypothetical examples, consider the following interchange, which took place as part of a conversation in 1998, in which I was inquiring about the location of various family groups prior to their settlement in Montetoni, the village in which this conversation took place.<sup>3</sup>

- (1) a. L: *Tyara pinejaxeri yonatan, pirijasantenixu o tinpijaxu?*  
 where you.saw.him yonatan at.Pirijasanteni or at.Tinpija

'Where did you see Yonatan, at Pirijasanteni or at Tinpija?'

- b. B: *Te, tera noneje janta. Ari axa xara pajirani noxemaxoti.*  
 not not I.see there indeed there there long.ago I.heard.about  
*Ixanti maixa ainyo janta nonej maixa. Maixa ixanti maixa*  
 he.said now there.is there I.saw now now he.said now  
*ainyo janta nonej maixa. Maixa paira tetyara pairani*  
 there.is there I.saw now now long.ago not.yet.then long.ago  
*iryo janta jatatsi ixanti axa ainyo axa tsinxateni.*  
 he there was.gone he.said there there.is there Tsinxateni

<sup>3</sup> Square brackets in the free translations below indicate systematically recoverable referents not present in the original Nanti, while braces indicate contextual information available only through the specific interaction.

*Speech Reporting Practices and Subjectivities*

‘I did not, did not see him there. Indeed, long ago there {at Tinpija} I heard about [him], he said, “There is someone there, I saw [him] {at Tsinxateni}.” At that time, long ago, before he {Yonatan} had left, he said, “There is someone there at Tsinxat-  
eni.”’

The speech reported in this data sample was produced approximately 25 years ago, and even under these circumstances, that speech is reported as direct speech.

A detailed examination of DSR in Nanti discourse reveals several patterns of use. First, Nanti speakers employ DSR, and only DSR, in contexts in which speakers of European languages regularly use performative verbs of saying, such as “promise”, “order”, and “prohibit”, in conjunction with indirectly reported speech. In short, Nanti speakers communicate about illocutionary forces deployed in conversation by directly reporting the speech carrying that illocutionary force.

Consider the following conversation, which took place in 1998, in which Tejotoro, a young man who is rising in political prominence in Montetoni, discusses with my partner and myself the conditions under which our presence in Montetoni is welcome:

- (2) T: *Tera, te tera nonxante, nonxante pijataje. Tera nonxante.*  
 not not not not.I.say not.I.say you.go.back not not.I.say  
*Oxanyota xatinxa onti pixa janta nojataje notimira.*  
 it.is.like noon it.is you.said there I.go.back my.land  
*Noxa nani, nani xameti. Tera nomintiganxajenpi noxa pijataje.*  
 I.say fine fine good not not.I.send.you.away I.say you.go.back  
*Jame nonejaxoti xanyorira saburi, xotsiro, jame nonejaxoti*  
 had.not I.see.about for.example machete knife had.not I.see.about  
*nonxa pijataje...*  
 I.will.say go.back

‘I don’t, don’t, don’t say, don’t say, “Go back” {i.e. go away}, I don’t say [it]. It is like you said at noon, “I am going back to where I live.” I say, “Fine, fine, good.” I don’t send you away, I don’t say, “Go back.” Had I not seen {that you brought}, for example, machetes and knives, had I not seen [them], I would say, “Go back”...’

Tejotoro continues:

- (3) T: *Itya nonejaxenpi tera tera nonxante janta mabani pisaninxaxa*  
 when I.see.you not not not.I.say there several your.people  
*pitentajigaxiti. Tera nonxante. Biro onti oga pixoriti*  
 you.accompany.over.here not not.I.say you it.is that your.wife  
*pintsipajigax. Noxanti jame jame pitentiro paniro*  
 you.will.accompany I.say had.not had.not you.accompany.her one  
*papuntaxa inxa iro yonta intinxami peresetente inxa*  
 you.came.alone he.will.say it that.one chief president he.will.say  
*pijataje.*  
 you.go.back

‘When I see you, I don’t, don’t say, “Several of your people [should] accompany you over here.” It is that one, your spouse, that you will accompany, I do not say [that] {i.e. that I should arrive with any other of “my people” apart from my spouse}. I say had you not, had you not come here with your spouse, had you come by yourself, he would say, that one, the chief, the president, he would say, “Go back.”’

This interchange is replete with instances of what speakers of European languages would likely characterize as “prohibiting”, “ordering”, or “demanding”, and which would likely be reported using a performative verb of saying and indirect speech. Instead, Tejotoro makes exclusive use of DSR to convey these illocutionary forces, as is typical in Nanti discourse.

Consider another example, in which Manoero, the present leader of Montetoni, describes the climax and resolution of the first friendly encounter with the Matsigenka, the neighboring indigenous group, in the late 1980s. At this point in the narrative, Manoero approaches the Nanti man who had gone to check on a Matsigenka man spotted gathering palm fronds by a Nanti hunting party.

- (4) M: *Inpo nopoxapajira. Ixanti maixa nonejajigaxiri nonejajigaxiri.*  
 then I.came.towards.there he.said now we.saw.him we.saw.him  
*Inpo ixanti, ixanti xamani pamutaxojigaxiri igapasite.*  
 then he.said he.said tomorrow you:PL.will.help.him his.palm.fronds  
*Inpo nojajigaxi. Patiro xutagite nojajigaxi. Xantira pamutaxoji*  
 then we.went one day we.went said.then you.all.help  
*inti matsigenxa. Ixanti te tsaxopi. ... Ixanti injax ixanti*  
 he.is person he.said not arrow he.said he.spoke he.said  
*maixa oxarioxa xapasi pamutaxojigaxenara. Nojajigaxitira.*  
 now that.one palm.fronde you.will.help.me we.went.over.there  
*Axa noxarajigaxi yonta ne, tobaini. noxarajiga axa, axa*  
 here we.measured this.one see many we.measured here here  
*noxarajiga. Inpo nojajig namugaxeni. Pasi yogatuti, pasi yogat.*  
 we.measured then we.went we.helped more he.got more he.got  
*Axa yopijotaxeni xapasi ojojojoi ... Yamataxotanara sintipoja.*  
 here he.piled.up palm.fronde wow he.float.on.away raft  
*Ixanti maixa nonpoxajira aixiro. Ixanti pamutaxojiga.*  
 he.said now I.will.come.back again he.said you:PL.will.help  
*Patiro aixiro sirijaga poxapaji aixiro.*  
 one again dry.season come.back again

‘Then I came back towards him. He said, “Now we saw him, we saw him” {i.e. the Matsigenka man}. Then he said, he said, “Tomorrow you all will help him with his palm fronds.” Then we went for one day, we went. He said, “No arrows” {i.e. he is not armed}. He said, “He spoke, he said, ‘Please you all help me with these palm fronds’, you will help him, he is a person” {i.e. as opposed to a violent killer}. We went, we numbered this many {gesturing}, this, see, we numbered many. This, this many {gesturing} we numbered. Then we went, we helped. More he took, more he took. He piled the palm fronds up, wow! He took [the palm fronds] away with a balsa raft. He said, “I will come back again.” He said, “Please help me [again]. In one more year, [I] will come back.”’

*Speech Reporting Practices and Subjectivities*

We see that Manoero makes exclusive use of DSR to communicate what European language speakers would be inclined to call “requests” and “promises”, terms they would typically incorporate as performative verbs that would introduce an indirect speech report.

Another relevant major use of DSR is the use of DSR by speakers of Nanti in contexts where speakers of European languages typically employ verbs of mental activity and mental state, such as “decide”, “believe”, and “think”. Consider the following example, in which Manoero describes the problems associated with an individual who was the nominal health worker in charge of health care in Montetoni:

- (5) M: *Pine mai xanta pitaxeni sanitari xara. Iriro janta mujigaxi,*  
you.see now stay stay.for health.worker there him there help.them  
*pine maixa iriro hanta, pine? Ari onxanta maixa*  
you.see now him there you.see indeed it.will.happen now  
*pijatajirixa jara ixanti totata nonxamosotaxitirira,*  
when.you.go.back will.not he.say wait! I.will.visit.him.over.there  
*ainyoxa otomi pine ainyoxa maixa isijajigata*  
there.is.perhaps her.son you.see there.is.perhaps now they.have.diarrhea  
*ainyoxa tyaxa ixantaxa janta imantsigajigax itomi.*  
there.is.perhaps whatever happened there they.are.sick his.son  
*Tya ixan? Jara ixant!*  
what he.says will.not he.say

‘You see, now he remains {in Maranxejari}, the health worker stays there for their benefit there, he helps them there, you see? Indeed, it will happen when you leave that he will not say, “Hold on! I am going to visit him over there, perhaps there is, perhaps there is a child, perhaps they [children] have diarrhea, perhaps, there is something going on there, sick children.” What does he say? He will not say [that]!’

What most speakers of European languages would describe as an attitude, a disposition, or perhaps an absence of a disposition—namely, disinterest on the part of the health worker to care for sick people in Montetoni—is instead rendered by Manoero as an absence of an utterance. Recall also Tejotoro’s discussion above, in which he expresses several opinions, evaluative judgments, and attitudes exclusively through reported speech. In that discussion, he even frames evaluative positions he is enunciating at that very moment as speech, thus concurrently reporting the speech he is producing.

Now consider a final example, in which Manoero describes the circumstances under which he came to visit his sick brother, who is the leader of the nearby Nanti community of Maranxejari. This reported exchange begins with the report by Manoero’s brother-in-law that Manoero’s brother is sick.

- (6) M: *Ixanti pirenti imantsigat. Irota nojatasita*  
 he.said your.brother he.is.sick it.agree I.went.purpose  
*noxamosotaxisitirira, noxa atsi nonxamosotaxite, je.*  
 I.visit.him.there.over.there I.said all.right I.will.visit.over.there yes

‘He said, “Your brother is sick.” For that reason I went to visit him over there. I said, “All right, I will visit over there, yes.”’

In this sequence, Manoero relates what a European language speaker would probably describe as his *decision* to visit his sick brother. As this example shows, however, Manoero does not characterize this process as “decision-making”, but rather uses DSR to report on his utterance announcing his departure.

### 3. SRPs and the Discursive Model of the Person

A pattern can be discerned among the preceding examples. Nanti speakers employ DSR to communicate about illocutionary forces, their own evaluative judgments or those of others, and expressions of agency arising in communicative interaction. From a communicative-functional perspective, then, DSR is a discursive practice which is used to communicate about individuals as perceiving and agentive social subjects,<sup>4</sup> and which is used to express in discourse the subjective orientations of these subjects towards social circumstances.

This use of DSR forms part of a broader pattern in Nanti discursive practice, in which the behavior of individuals is discussed in terms of the utterances they produce and the physical actions they perform. My detailed examinations of recordings of Nanti discourse have not yet revealed cases of speculation about what someone thought, what someone intended, what they meant, or what they decided.<sup>5</sup> Instead, I have found minute discussions of what people *said* and equally careful discussions of what people did *not* say, which frequently appears to be as relevant as what people *did* say. Thus, not only is DSR employed to discursively represent the subjectivity and agency of individuals, but, apart from descriptions of physical actions, DSR is the *principal* (if not exclusive) means for doing so.

As indicated above, the *internal states* of individuals are not frequent or central topics of discourse, nor are internal states a significant explanatory or analytic modality in discourse for assessing or understanding the behavior of individuals. Nanti individuals thus appear to organize their communication about agency and

---

<sup>4</sup> The perceptive reader will have noted that not all occurrences of DSR in the data correspond to the uses just outlined. Another important aspect of the use of DSR, not discussed in this paper for reasons of space, relates to Nanti epistemology and issues of evidentiality and representations of experience in Nanti discourse. Nanti speakers tend to be circumspect in their knowledge claims, and typically prefer to relate by means of reported speech knowledge that they have acquired solely through talk (Michael 2001).

<sup>5</sup> This does not, of course, rule out the possibility of such speculation or discussion in Nanti discourse. It does point, however, to its rarity—either because of the lack of salience of such a viewpoint on agency and subjectivity, or because of its sensitivity (cf. Besnier 1993).

## *Speech Reporting Practices and Subjectivities*

subjectivity by employing a model of the person in which the locus of subjectivity and agency is speech, rather than, say, mental or intentional states.

We can characterize the model of the person operative in Nanti discourse as a *monistic* model. Crucially, a monistic model of the person takes speech and the meaningful states and activity of the social subject to be inseparably fused. Consonant with such a model is the notion that speech is not a 'mere' conduit for the projection of basic internal states, but rather that the utterance is the basis of meaning. This contrasts with a Cartesian, or dualistic, conception of the person, in which meaning arises from mental states (Vendler 1972).

It should be noted that I am making a claim about the *discursive model* of the person employed by Nanti speakers, a model of the person that Nanti speakers use to organize their *talk* about the behavior of individuals. How one talks about a topic and how one might otherwise relate to that topic are logically distinct, and I am leaving open for the moment the question of whether this model of the person further penetrates other areas of Nanti activity and society.

Further evidence for the use of a monistic discursive model of the person by Nanti speakers comes from the observed exclusive use of DSR to report speech. As Alan Rumsey has pointed out, indirect speech reporting practices depend critically on the distinction between utterance and meaning (Rumsey 1990:347). This distinction, Rumsey argues, is based on a linguistic ideology that takes utterances to encoding pre-existent 'mental meanings'. In terms of this ideology, ISR 'extracts' meaning from the original utterance, thereby allowing the speaker to efficiently strip the 'content' from the 'code'.

If, however, speakers employ a monistic discursive model of the person, the utterance/meaning distinction is problematic. Specifically, a monistic model excludes the possibility of pre-existent internal meanings that are subsequently encoded in messages, since a monistic model makes no place for the existence of meanings as distinct from utterances. Consequently, the notion of extracting meaning from an utterance, which is basic to ISR, is incompatible with a monistic model of the person. We would therefore expect that speakers employing a monistic model of the person would report speech directly, rather than indirectly, since DSR does not require the splitting apart of 'meaning' and utterance. This is, of course, precisely the situation I have observed among Nanti speakers.

Thus, by positing that a monistic model of the person underlies Nanti speaking practices, it is possible to account for both the exclusive use of DSR for the reporting of speech, and for the use of DSR for communicating about the agency and subjectivity of individuals. The preceding analysis therefore suggests that DSR-dominant speaking practices are a coherent set practices of representation of the social subject, and that the coherence of these speaking practices is not coincidental, but stems from their grounding in a monistic discursive model of the person.

#### 4. Comparative Evidence

Evidence from other societies in lowland Latin America and elsewhere suggests that the characteristics and patterning of Nanti SRPs are not an isolated phenomenon.

Ellen Basso's areally related work on narrative and biography among the Kalapalo, an indigenous group of the Xingu River area of central Brazil, suggests that Kalapalo speakers employ reported speech in a manner similar to that found in Nanti society. Basso remarks that speech is normally directly reported by Kalapalo speakers, and that, in addition, when Kalapalo speakers relate personal and historical narratives, they reveal the motives, emotions, and the subjective interpretations of reality of the characters in these narratives through reported speech, rather than through a description of mental or emotional states (Basso 1995:295-296).

Likewise, Joel Sherzer, describing the speaking practices of the Kuna of Panama, indicates that DSR is frequently used in place of representations of thought introduced by mental activity verbs, and that the meaning/utterance distinction described above is considerably less salient in Kuna society than in Euro-American society (Sherzer 1983). As is the case with Nanti society, Sherzer reports that the direct, rather than indirect, reporting of speech is the dominant means for reporting the utterances of others.

Shifting away from indigenous Latin America, Alan Rumsey, in his work on the language ideological dimensions of reported speech in Ungarinyin, a language spoken in northwestern Australia, indicates that no distinction is drawn between indirect and direct speech in Ungarinyin discourse (Rumsey 1990:347). At the same time, Rumsey argues, Ungarinyin speakers do not appear to find salient the distinction between 'meaning' and 'wording' (ibid.:354), a distinction already noted to be basic to the dualistic model of the person.

At present, the sparseness of ethnographic data focusing on the communicative-functional character of SRPs unfortunately leaves many important questions unanswered. For example, the cultural and geographic distribution of monistic models of the person, the ways in which monistic models of the person are implicated in the organization of speaking practices in particular societies, and the variations in the constitution of these models remain unknown.

There is tantalizing evidence that suggests that the use of a monistic model of the person in discourse, indicated by the co-occurrence of DSR-dominant SRPs and use of DSR to represent agency and subjectivity in discourse, is a widespread areal discourse trait in lowland Latin America (Waltz 1976, Witte 1976). Similarly, Rumsey indicates that the linguistic ideology he described among the Ngarinyin is shared by other Australian Aboriginal groups (Rumsey 1990:352-345). Answers to these questions of distribution and variation must, however, await further ethnographic research.

Despite the current paucity of comparative evidence, we see indications that particular speaking practices co-occur with particular models of the person. On the one hand, we find European societies exhibiting ISR-centered SRPs and a

Cartesian model of the person in which dualist understandings of mind and body, talk and thought, and speech and meaning hold sway, while on the other hand we have Nanti, and probably Kalapalo, Kuna, and Ungarinyin as well, displaying DSR-centered SRPs and a model of the person in which the dualisms that constitute the Cartesian subject are either much less salient, or absent entirely.<sup>6</sup>

### **5. SRPs and Forms of Subjectivity**

These observations cast an interesting light on recent work on the relationship between speaking practices and formulations of subjectivity. In particular, the observed features of SRPs in Nanti society (and possibly other indigenous societies in Latin America and Australia) provides complementary evidence for the claims of Benjamin Lee that the particular Western form of subjectivity known as Cartesian dualism is based on grammatical analogies between speaking and thinking inherent in European speaking practices, especially SRPs (Lee 1997).

Lee develops his argument by showing that the grammatical properties of indirect speech and its interpretation, and the gradation of subjective expressiveness in performative verbs and verbs of thinking, form the semiotic basis for a 'subjective realm' that leads, by metaphorical transference, to notions of an internal 'mental' realm, and of 'mental meaning', which are further reified in the system of performative verbs and mental state and activity verbs found in European languages (ibid.:202-221).

While Lee's subtle and well-supported argument is compelling in many respects, it is based entirely on the internal structures of European languages and speaking practices, and brings no comparative data to bear on the issue of the relationship between SRPs and subjectivities. The possibility arises, then, that the co-occurrence in European societies of a Cartesian model of subjectivity and ISR-dominant speaking practices is simply coincidental, despite the apparent semiotic analogies between the organization of speaking practices and models of the speaking subject.

However, the evidence from Nanti society suggests that SRPs and forms of subjectivity are meaningfully related to one another. Lee argues that Cartesian subjectivity is grounded in the use of performatives, mental activity verbs, and ISR practices found in European societies. On the basis of this, we would expect that in a society in which these speaking practices are absent, we would find a rather different formulation of subjectivity. In particular, we would expect the absence of a dualistic formulation of subjectivity.

---

<sup>6</sup> It is probably the case that in any system of discursive practice, elements of both discursive models that I have described are present to some degree. After all, even in American English communicative settings, disavowal of racist *intent* in producing an utterance that someone finds racist goes only so far in mitigating the effect of the utterance. Similarly, arguing that a promise one gave did not constitute actually a promise, because of a lack, when uttering the promise, of a concomitant intent to fulfill the promise (a violation of the felicity conditions for the speech act, cf. Austin 1962), is hardly compelling.

This, plausibly, is what we find in Nanti society. Nanti speaking practices eschew performatives; mental activity verbs are rare or absent;<sup>7</sup> and DSR is the exclusive form of speech reporting. At the same time, the model of the person upon which Nanti speaking practices are based is a non-dualistic, monistic one, in which meaning/utterance, thought/action, and mind/body dualisms are not salient.

In short, the evidence from Nanti speaking practices provides comparative support for Lee's claims that Cartesian subjectivity is grounded in particular speaking practices. At the same time, Lee's work on the discursive basis of Cartesian subjectivity provides a theoretical basis for understanding the interdependence of speaking practices and forms of subjectivity.

It should be noted, though, that in my arguments above I am making a more modest claim than that advanced by Lee. While Lee relates speaking practices to *forms of subjectivity*, I am relating speaking practices to discursive models of the person—that is, to a system of *representations of subjectivity* in discourse.

Whether the monistic model of the person that I claim is employed by Nanti speakers is solely a discursive model that they use to organize speaking, or whether it penetrates into other realms of action, cognition, and interaction, is just one of several open questions that remain. While the widespread nature of Cartesian subjectivity in Western societies is amply attested, whether the monistic model of the person that appears to be operative in the organization of Nanti speaking practices is an extension of more general Nanti conceptions of the person and the social subject, or whether it is a model that serves solely to organize discursive practices, remains an open question.<sup>8</sup> It is conceivable, for example, that while reference to 'internal' states or actions is rigorously avoided in Nanti discourse, Nanti individuals otherwise do employ a dualistic model of the person in understanding the behavior of individuals.<sup>9</sup> No doubt methodologies developed by cognitive and developmental psychologists concerned with theories of mind could be profitably adapted to explore these issues.

Part of the difficulty in determining the extent of the use of this monistic model in the ideological and cognitive organization of Nanti society arises, I feel,

---

<sup>7</sup> Nominal forms corresponding to the mentalistic terms found in European languages such as "knowledge", "opinion", or "belief" appear to be absent in Nanti. A small number of verb stems may admit dualistic or mentalistic interpretations, but monistic interpretations of these stems centering on speech and action, rather than on 'internal' processes, are equally compelling.

<sup>8</sup> A prosaic example of a model employed in Euro-American discourse that does not extend fully into the cognitive realm is the ANGER IS HEAT metaphor (Lakoff 1986:382-387). This metaphor is employed in a highly productive manner in talking about anger, yet Americans do not believe that dumping cold water on an angry person, or putting such a person in a refrigerator, is an effective means for diminishing their anger. The ANGER IS HEAT metaphor is certainly a discursive model, but it only partially penetrates the American cultural etiology of anger.

<sup>9</sup> Such a situation is described by Niko Besnier in his study of reported speech and affect on Nukulaelae (Besnier 1993). Besnier indicates that Nukulaelae islanders consider speculation or inference about the internal states or thoughts of others to be "irresponsible" and "demented", although such activity is attributed to "children, adolescents, and irresponsible gossipers" (ibid.:166).

from difficulty in conceiving of what a monistic conception of the person might constitute. How, in short, do we theorize a conception of the subject that does not incorporate either mental meaning or mental activity as separate from speech and action? Alessandro Duranti's work on Samoan ethnopragmatics suggests some directions, but much work remains to be done if we are to develop an analytically robust notion of the monistic subject (Duranti 1992).<sup>10</sup>

## **6. Conclusions**

This paper is intended to contribute to our ethnographic understanding of reported speech. In particular, I have attempted to account for the communicative-functional properties of reported speech in Nanti society, and for the exclusive use of DSR to report speech, by appealing to the notion that Nanti speakers organize their representations of subjectivity and agency in discourse in terms of a specific model of the person. I have characterized this model as a *monistic* model of the person, in which meaning and utterance are taken to be fused, and which concomitantly does not recognize dualistic distinctions between mind and body, and thought and action. Whether this model is solely employed to organize talk, or if it also organizes other aspects of cognition and behavior remains a question for future work.

## **References**

- Aikhenvald, Alexandra. 1999. The Arawak Language Family. In R.M.W. Dixon and Alexandra Aikhenvald (eds.), *The Amazonian Languages*. New York: Cambridge University Press.
- Austin, J. L. 1962. *How to Do Things with Words*. Cambridge: Harvard University Press.
- Basso, Ellen. 1995. *The Last Cannibals: A South American Oral History*. Austin, TX: University of Texas Press.

---

<sup>10</sup> This stands in sharp contrast with dualistic models of the person, which are well-developed in both their folk-theoretic forms and as formal models derived from folk-theoretic forms. The dualistic model of the person employs a vessel metaphor for the person (cf. Lakoff 1987:381-388), in which the body serves as a container for a medium (mind) in which meaning is produced, sustained, and consumed. Meaning is transferred from the interior of one vessel (person) to another through words, which serve as a conduit for the meaning (thoughts) contained in the mind of the speaker (Reddy 1979).

In the context of this familiar model, indirect speech reporting is understood to be the process of extracting and representing the meaning contained in speech, which originates in the internal, mental realm of the speaker. Speech acts are distinguishable from the utterances themselves which a speaker produces when performing them by the felicity condition that the speaker hold particular mental states in order that the utterance count as a particular speech act (Austin 1962:8-11). Similarly, mental activity takes place in the inner medium of the mind and manifests itself as acts of agency and subjective perceptions of the surrounding physical and social world.

Lev Michael

- Besnier, Niko. 1993. Reported Speech and Affect on Nukulaelae Atoll. In Jane Hill and Judith Irvine (eds.), *Responsibility and Evidence in Oral Discourse*. New York: Cambridge University Press, 161-181.
- Duranti, Alessandro. 1993. Intentions, Self, and Responsibility: An Essay in Samoan Ethnopragmatics. In Jane Hill and Judith Irvine (eds.), *Responsibility and Evidence in Oral Discourse*. New York: Cambridge University Press, 24-47.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Lee, Benjamin. 1997. *Talking Heads: Language, Metalanguage, and the Semiotics of Subjectivity*. Durham, NC: Duke University Press.
- Michael, Lev. 2001. Knowledge, Experience, and Reported Speech in an Amazonian Society: The Nanti of Southeastern Peru. *Proceedings of the Ninth Symposium about Language and Society – Austin*. Austin, TX: Texas Linguistic Forum.
- Reddy, Michael. 1979. The Conduit Metaphor. In A. Ortony (ed.), *Metaphor and Thought*. New York: Cambridge University Press.
- Rumsey, Alan. 1990. Wording, Meaning, and Language Ideology. *American Anthropologist* 92:246-361.
- Sherzer, Joel. 1983. *Kuna Ways of Speaking: An Ethnographic Perspective*. Austin, TX: University of Texas Press.
- Vendler, Zeno. 1972. *Res Cogitans: An Essay in Rational Psychology*. Ithaca, NY: Cornell University Press.
- Waltz, Nathan. 1976. Discourse Functions of Guanano Sentence and Paragraph. In Robert Longacre (ed.), *Discourse Grammar: Studies in Indigenous Languages of Colombia, Panama, and Ecuador*. Dallas: Summer Institute of Linguistics, 21-145.
- Witte, Paul. 1976. Functions of the Andoke Copulative in Discourse and Sentence Structure. In Robert Longacre (ed.), *Discourse Grammar: Studies in Indigenous Languages of Colombia, Panama, and Ecuador*. Dallas: Summer Institute of Linguistics, 253-288.

Lev Michael  
Department of Anthropology  
University of Texas at Austin  
Austin, TX 78712-1086

lmichael@mail.utexas.edu

## Lexical Tone and Markedness in Standard Thai

BRUCE MORÉN<sup>1</sup> and ELIZABETH ZSIGA<sup>2</sup>  
*University of Tromsø*<sup>1</sup> and *Georgetown University*<sup>2</sup>

### 1. Data

Every first-year linguistics student knows the five different ways to say *naa* in Thai. He or she also knows that an autosegmental representation, with primitives high (H) and low (L), represents the contrasts between the mid, high, low, falling and rising tones both simply and elegantly, as shown in (1). A closer look at the distribution of these five tones, however, reveals that the system is not so simple after all.

|     |              |                                                 |                                                 |                                                                         |                                                                         |
|-----|--------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------------------------------|-------------------------------------------------------------------------|
| (1) | <u>mid</u>   | <u>high</u>                                     | <u>low</u>                                      | <u>falling</u>                                                          | <u>rising</u>                                                           |
|     | $\sigma$     | $\begin{array}{c} H \\   \\ \sigma \end{array}$ | $\begin{array}{c} L \\   \\ \sigma \end{array}$ | $\begin{array}{c} H \quad L \\ \diagdown \quad / \\ \sigma \end{array}$ | $\begin{array}{c} L \quad H \\ / \quad \diagdown \\ \sigma \end{array}$ |
|     | [ná:]        | [ná:]                                           | [nà:]                                           | [nâ:]                                                                   | [nǎ:]                                                                   |
|     | ‘rice field’ | ‘custard apple’                                 | ‘aunt’                                          | ‘face’                                                                  | ‘thick’                                                                 |

First, there is an interaction of tone with stress. Tone is realized only on stressed syllables: in unstressed syllables all tones reduce to mid.<sup>1</sup> So, for example, in the reduplicated form [sawsǎ:w] ‘young girls’ the underlying rising tone on /sǎ:w/ is realized only on the stressed final syllable, while the initial syllable is pronounced with a mid tone. This interaction, we will argue, is a straightforward example of positional faithfulness.

Other restrictions are more complex and puzzling. Not all stressed syllables can bear all tones. Only open syllables with long vowels (CVV), or any syllable closed by a sonorant, regardless of vowel length (CVS and CVVS), can bear all five tones. Examples are given in (2).

---

<sup>1</sup> There are counterexamples to this claim, but we follow Yip’s (1982:88) suggestion that these seem to involve morphology and secondary stress effects that we will not address here.

|     |         |                       |  |                                |  |                                |
|-----|---------|-----------------------|--|--------------------------------|--|--------------------------------|
| (2) |         | <u>CVV</u>            |  | <u>CVS</u>                     |  | <u>CVVS</u>                    |
|     | mid     | [na:] ‘rice field’    |  | [daŋ] ‘loud’                   |  | [t <sup>h</sup> a:ŋ] ‘way’     |
|     | low     | [nâ:] ‘custard apple’ |  | [sàŋ] ‘to order’               |  | [p <sup>h</sup> à:n] ‘to pass’ |
|     | high    | [ná:] ‘aunt’          |  | [sám] ‘to repeat’              |  | [sá:j] ‘left’                  |
|     | falling | [nâ:] ‘face’          |  | [nân] ‘that’                   |  | [hâ:m] ‘to prohibit’           |
|     | rising  | [nǎ:] ‘think’         |  | [k <sup>h</sup> ǎm] ‘to order’ |  | [hǎ:n] ‘to divide’             |

Stressed open syllables with short vowels (CV) do not occur. CV is allowed only in unstressed syllables, where it has mid tone.

On syllables closed by obstruents, the tonal types are restricted in what appears to be a very odd way. On CVO syllables, only high and low can occur. On CVVO syllables, only falling and low are allowed.<sup>2</sup> Examples are given in (3).

|     |         |                  |  |                             |
|-----|---------|------------------|--|-----------------------------|
| (3) |         | <u>CVO</u>       |  | <u>CVVO</u>                 |
|     | mid     | ----             |  | ----                        |
|     | low     | [càp] ‘catch’    |  | [k <sup>h</sup> à:t] ‘torn’ |
|     | high    | [lák] ‘to steal’ |  | ----                        |
|     | falling | ----             |  | [mâ:k] ‘many’               |
|     | rising  | ----             |  | ----                        |

The absence of mid tone on obstruent-final syllables creates a markedness paradox. If mid is the lack of a phonological tone specification (a fairly standard assumption), and lack of tone is less marked than the presence of tone, why is a tone required on these syllables? The lack of high on CVVO is also a markedness paradox, since falling (a contour tone) should imply high (a simple tone). In addition, one must explain the absence of rising tone on CVVO and the lack of any contours at all on CVO.

These distributional gaps have been discussed in previous literature (Kruatrachue 1960, Abramson 1962, Gandour 1974, Yip 1982), but no explanation has been proposed. Regarding the lack of mid tone on obstruent-final syllables, Yip (1982:89) says that “the absence of M is odd, and neither Gandour or I have any explanation to offer...” In this paper, we propose an analysis of these surprising tonal patterns, drawing on principles of both phonetics and phonology.

<sup>2</sup> According to Gandour (1974), the rare exceptions with CVVC-H (e.g. [k<sup>h</sup>á:t] ‘card’) and CVC-HL (e.g. [k<sup>h</sup>lák] ‘crowded’) are either loan words or onomatopoeia. Yip (1982) disputes this claim, but does not provide much data. We leave the analysis of these exceptions to future research.

**2. Stress and Tone**

Within an OT analysis, the iambic stress pattern of Thai is easily captured by assuming a high ranking constraint ALIGN<sub>HD-R</sub> that requires the head syllable of the foot (the stressed one) to fall at the right edge of the foot. The fact that tonal contrasts are realized only on stressed syllables is a straightforward example of positional faithfulness (Beckman 1995, Alderete 1995, Morén 1999). Underlying specifications surface in prominent positions, but are neutralized in non-prominent positions. We assume a positional faithfulness constraint MAX<sub>STRESS</sub>[T], which requires realization of underlying tones in stressed syllables. This positional faithfulness constraint outranks the markedness constraint against having a tone, \*[T]. On the other hand, the general faithfulness constraint MAX[T], which calls for no loss of tone in *any* context, is low-ranked. As shown in (7), the ranking Positional Faithfulness » Markedness » General Faithfulness ensures that tone is realized on stressed syllables, but underlying tonal specifications do *not* surface on unstressed syllables. Syllables phonologically unspecified for tone are realized as mid.

- (4) MAX<sub>STRESS</sub>[T]: do not delete an underlying tone from a stressed syllable.
- (5) \*[T]: do not have tone.
- (6) MAX[T]: do not delete underlying tones.
- (7) Neutralization of tone in unstressed syllables

| /saw-LH + saw-LH/ | MAX <sub>STRESS</sub> [T] | *[T]  | MAX[T] |
|-------------------|---------------------------|-------|--------|
| saw + 'sa:w       | *!*                       |       | ****   |
| saw-LH + 'sa:w    | *!*                       | **    | **     |
| saw-LH + 'sa:w-LH |                           | ***!* |        |
| ☞ saw + 'sa:w-LH  |                           | **    | **     |

**3. Background: V-Length, C-Weight, and Glottal Epenthesis**

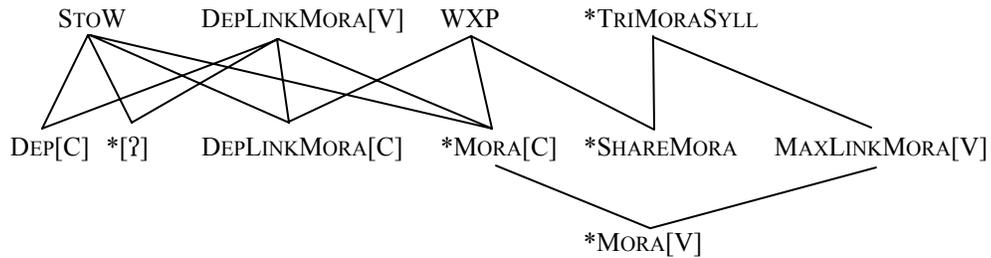
In order to account for the facts of vowel length and syllable weight in Thai, we will assume the constraints in (8)-(17), based on the analysis of moraic structure in Morén (1999). The proposed constraint ranking is given in (18).

*Markedness constraints:*

- (8) \*MORA[V]: do not associate a mora with a vowel.
- (9) \*MORA[C]: do not associate a mora with a consonant.
- (10) \*TRIMORASYLL: no tri-moraic syllables.
- (11) \*SHAREMORA: there should be only one segment associated with a mora.
- (12) STOW (stress-to-weight principle): stressed syllables must be heavy.
- (13) WXP (weight-by-position): syllable-final consonants should be moraic.
- (14) \*[ʔ]: glottal stops are prohibited.

*Faithfulness constraints:*

- (15) MAXLINKMORA[V]: do not delete an underlying mora from a vowel.<sup>3</sup>
- (16) DEPLINKMORA[V]: do not add a mora to a vowel that it did not have underlyingly.<sup>4</sup>
- (17) DEP[C]: do not insert a consonant.
- (18) Constraint rankings for length/weight/glottal epenthesis



In Thai, short (mono-moraic) and long (bi-moraic) vowels are contrastive, but there are no trimoraic syllables. To ensure that one or two moras associated to a vowel in the input will surface in the output, MAXLINKMORA[V] is ranked above \*MORA[V]. However, because \*TRIMORASYLL outranks MAXLINKMORA[V], a third mora linked to a vowel will not surface.

Open syllables with short vowels occur only in unstressed positions, so we assume that all stressed syllables in Thai must be bimoraic (STOW » \*MORA[V]). We further assume that all final consonants in Thai are moraic, because CVC syllables *can* bear stress (STOW » \*MORA[C]).

Underlyingly short, open syllables that occur in a stressed position are made heavy by the addition of a final glottal stop (STOW » DEP[C], \*[ʔ], \*MORA[C]). As shown in tableau (19), glottal stop insertion is preferred to vowel lengthening as the method for creating heavy syllables (DEPLINKMORA[V] » DEP[C], \*[ʔ], \*MORA[C]). We indicate moras associated with a segment by superscripting.

(19) Syllable weight augmentation via glottal epenthesis, not V-lengthening

| /CV <sup>μ</sup> /             | DEPLINKMORA[V] | STOW | DEP[C] | *[ʔ] | *MORA[C] |
|--------------------------------|----------------|------|--------|------|----------|
| CV <sup>μ</sup>                |                | *!   |        |      |          |
| CV <sup>μ</sup> ʔ              |                | *!   | *      | *    |          |
| CV <sup>μμ</sup>               | *!             |      |        |      |          |
| CV <sup>μ</sup> ʔ <sup>μ</sup> |                |      | *      | *    | *        |

<sup>3</sup> Morén (1999), simplified here.

<sup>4</sup> Morén (1999), simplified here.

We will argue that codas following long vowels share the second mora of the vowel, as shown in (20). Mora sharing is supported by the phonetic evidence: long vowels in closed syllables are shorter than in open syllables. Broselow et al. (1998) attribute vowel shortening in closed syllables to the fact that the following consonant “takes up” some of the time of the last mora. We will see in §5 that the tonal evidence from CVVO syllables also supports this conclusion.

(20) Mora sharing between long vowel and coda consonant

| /CV <sup>μμ</sup> C/                                                              | WXP | *TRIMORASYLL | *SHAREMORA | *MORA[C] |
|-----------------------------------------------------------------------------------|-----|--------------|------------|----------|
|  |     |              | *          | *        |
|  | *!  |              |            |          |
|  |     | *!           |            | *        |

#### 4. CVV, CVS, CVVS Syllables

We now turn to the analysis of tonal patterns in CVV, CVS, and CVVS syllables. When these syllables are stressed, mid, high, low, rising, and falling tones are permitted. As we saw in §2, the positional faithfulness constraint MAXSTRESS[T] outranks the markedness constraint \*[T], ensuring that the surface representation is faithful to the input. Tone metathesis is prohibited by LINEARITY (“do not switch linear order”). An example of a rising tone on a CVV syllable is given in (22).

In this section, examples using CVV syllables will be given, but the result would be the same for CVS and CVVS. At this point, we assume that \*[H] and \*[L], as well as MAXSTRESS[H], MAXSTRESS[L], and LINEARITY, are not ranked with respect to each other. We will argue for a specific ranking of the faithfulness constraints later.

(21) LINEARITY: no metathesis

(22) Contrastive rising tone in stressed CVV syllables

| /CVV-LH/                                                                            | MAXSTRESS[H] | MAXSTRESS[L] | LINEARITY | *[H] | *[L] |
|-------------------------------------------------------------------------------------|--------------|--------------|-----------|------|------|
|  |              |              |           | *    | *    |
| CVV-H                                                                               |              | *!           |           | *    |      |
| CVV-L                                                                               | *!           |              |           |      | *    |
| CVV                                                                                 | *!           | *!           |           |      |      |
| CVV-HL                                                                              |              |              | *!        |      |      |

We will assume that the mora is the tone-bearing unit. Association of H to one mora and L to another will produce rising and falling contours. More complex contours are ruled out by a constraint prohibiting the association of more than one tone to a single mora, as shown in (25).

(24) \*[TT]<sub>μ</sub>: no more than one tone per mora.

(25) Neutralization of complex contour tones

| /CVV-LHL/ | *[TT] <sub>μ</sub> | MAXSTRESS[H] | MAXSTRESS[L] | LINEARITY | *[H] | *[L] |
|-----------|--------------------|--------------|--------------|-----------|------|------|
| CVV-LHL   | *!                 |              |              |           | *    | **   |
| ☞ CVV-LH  |                    |              | *            |           | *    | *    |
| ☞ CVV-HL  |                    |              | *            |           | *    | *    |
| CVV-H     |                    |              | **!          |           | *    |      |
| CVV-L     |                    | *            | *!           |           |      | *    |

Note that if the input string has three (or more) tones, we don't know whether the output would be rising or falling, but we do know it would be a simple contour tone.

Mid tone (lack of tone) comes for free. If the input is unspecified for tone, the \*[H] and \*[L] markedness constraints will prefer an output without a tone (unless there is a highly-ranked constraint requiring that a tone be present).

(26) Contrastive mid tone in stressed CVV syllables

| /CVV/  | MAXSTRESS[H] | MAXSTRESS[L] | *[H] | *[L] |
|--------|--------------|--------------|------|------|
| CVV-HL |              |              | *!   | *    |
| CVV-H  |              |              | *!   |      |
| CVV-L  |              |              |      | *!   |
| ☞ CVV  |              |              |      |      |

## 5. CVVO Syllables

We now turn to the analysis of syllables closed by obstruents, beginning with those with long vowels. These syllables must be realized with either low or falling tone, not mid, high, or rising. The generalization seems clear: these syllables must end low. We propose to account for this generalization with a constraint that coda obstruents must be associated with low tone.

(27) OBSCODA → L: coda obstruents must be associated with L tone.

This constraint makes phonetic sense because, in Thai, coda obstruents are voiceless and glottalized. According to Maddieson (1977), “a simple cessation of voicing in [coda] position could readily be reinterpreted as a laryngeal segment.” As voicelessness is reinterpreted as phonological laryngealization, the laryngealization may in turn lead to increased vocal fold stiffness and slower vibration prior

to complete glottal closure. Abramson (1962) documents glottalization of final consonants in Thai, accompanied by low-frequency glottal pulses on the preceding vowel. These low-frequency pulses may give rise to the perception of low tone, and the phonologization of that percept is the constraint in (27). Such a progression is given further support by Diller (1996), who shows that unvoiced segments were associated with non-high tones in the development of Thai.

Certainly, the cross-linguistic relationship between tone and laryngeal features is complicated. Maddieson (1976, 1977), for example, cites three cases (Navajo, Kiowa, and Kapanahua) where glottalized codas are associated with low tone and four cases (Jeh, Vietnamese, Danish, and Latvian) where glottalized codas are associated with *high* tone. Maddieson specifically argues that the association between high tone and glottalization comes about when, in order to reach a high pitch at the end of a long syllable, the vocalis muscle is tensed beyond the parameters necessary to continue modal voicing. Extra glottal tension may also be recruited to extend voicing in low tones, however: a speaker’s voice can “crack” at both extremes of her or his range. Interestingly, an EMG study of two Thai speakers by Erickson and Abramson (1972) found active tensing of the vocalis muscle itself in only one context: the end of the falling tone.

We now turn to working out the phonological consequences of our proposed constraint. High ranking of  $\text{OBSCODA} \rightarrow \text{L}$  solves our first two markedness paradoxes. Regardless of the input tonal sequence, only low and falling tones will surface on CVVO syllables.

As shown in tableau (28), if the input is associated with a low tone, or if there is no underlying tonal specification, the syllable will surface with a low tone. Since our constraint is a positive one, *requiring* the presence of a low tone (not just, for example, prohibiting H in this position), mid tones, with no tonal specification at all, cannot surface on these syllables.

(28) **Markedness Paradox 1:** neutralization to low (not mid) in CVVO

|          |                         |      |      |
|----------|-------------------------|------|------|
| /CVVO-L/ | OBSCODA $\rightarrow$ L | *[H] | *[L] |
| ☞ CVVO-L |                         |      | *    |
| CVVO     | *!                      |      |      |
| /CVVO/   | OBSCODA $\rightarrow$ L | *[H] | *[L] |
| ☞ CVVO-L |                         |      | *    |
| CVVO     | *!                      |      |      |

Tableau (29) shows that if there is a high tone in the input, a falling tone will surface. High and rising both fatally violate  $\text{OBSCODA} \rightarrow \text{L}$ , while simple low unnecessarily violates  $\text{MAXSTRESS}[\text{H}]$ . The mid tone violates both.

(29) **Markedness Paradox 2:** neutralization to falling (not high) in CVVO

| /CVVO-H/ | OBSCODA→L | MAXSTRESS[H] | *[H] | *[L] |
|----------|-----------|--------------|------|------|
| CVVO-L   |           | *!           |      | *    |
| CVVO-H   | *!        |              | *    |      |
| CVVO-LH  | *!        |              |      |      |
| CVVO-HL  |           |              | *    | *    |
| CVVO     | *!        | *            |      |      |

LINEARITY, as shown in (30), must be ranked fairly low in the hierarchy. Both falling and rising input sequences will surface as falling.

(30) **Tone metathesis in CVVO**

| /CVVO-HL/ | OBSCODA→L | MAXSTRESS[H] | LINEARITY | *[H] | *[L] |
|-----------|-----------|--------------|-----------|------|------|
| CVVO-L    |           | *!           |           |      | *    |
| CVVO-H    | *!        |              |           | *    |      |
| CVVO-LH   | *!        |              | *         |      |      |
| CVVO-HL   |           |              |           | *    | *    |
| CVVO      | *!*       |              |           |      |      |

| /CVVO-LH/ | OBSCODA→L | MAXSTRESS[H] | LINEARITY | *[H] | *[L] |
|-----------|-----------|--------------|-----------|------|------|
| CVVO-L    |           | *!           |           |      | *    |
| CVVO-H    | *!        |              |           | *    |      |
| CVVO-LH   | *!        |              | *         |      |      |
| CVVO-HL   |           |              | *         | *    | *    |
| CVVO      | *!*       |              |           |      |      |

Thus, no matter what the tonal specification of the input, high ranking of OBSCODA→L ensures that the output in CVVO syllables will have either a low or falling tonal pattern.

**6. CVO Syllables**

Finally, we turn to the CVO syllables, where the only tones allowed are simple H and L. These syllables have the inherent difficulty that voiceless obstruents cannot realize tone. With no vibration of the vocal folds, there cannot be a tone. Therefore, tones must be realized on the monomoraic vowel. We will formalize this with the constraint in (31).

(31) **REALIZETONE:** tones must be associated to a segment that can support vocal fold vibration.

In CVVO syllables, this constraint presents no problem. OBSCODA→L requires that a low tone be associated to the mora linked to the final consonant. But since this mora is shared with a vowel, the tone can be realized, and a low or falling specification results, as shown above. In CVO syllables, however, though

*Lexical Tone and Markedness in Standard Thai*

two morae are present, only one is able to realize a tone. Given the constraint against more than one tone linked to a single mora, only simple tones can occur.

If there is a low tone, or no tone, in the input, the output will be a low tone linked to both moras. This violates low-ranked  $*[\mu\mu]T$ , which penalizes association of more than one mora to a single tone, but crucially satisfies both  $OBSCODA \rightarrow L$  and  $REALIZETONE$ .

(32) Tone neutralization to low in CVO with tone shared by moras

| /CVO-L/                                                                   | REALIZETONE | OBSCODA→L | *[L] | *[μμ]T |
|---------------------------------------------------------------------------|-------------|-----------|------|--------|
| $\begin{array}{c} \mu \mu \\ \text{CVO} \end{array}$                      |             | *!        |      |        |
| $\begin{array}{c} L \\   \\ \mu \mu \\ \text{CVO} \end{array}$            | *!          |           | *    |        |
| $\begin{array}{c} L \\ / \backslash \\ \mu \mu \\ \text{CVO} \end{array}$ |             |           | *    | *      |
| /CVO/                                                                     | REALIZETONE | OBSCODA→L | *[L] | *[μμ]T |
| $\begin{array}{c} \mu \mu \\ \text{CVO} \end{array}$                      |             | *!        |      |        |
| $\begin{array}{c} L \\   \\ \mu \mu \\ \text{CVO} \end{array}$            | *!          |           | *    |        |
| $\begin{array}{c} L \\ / \backslash \\ \mu \mu \\ \text{CVO} \end{array}$ |             |           | *    | *      |

Yet high tones can also occur on CVO syllables. If there is a high tone in the input, it must surface. Tableau (33) shows that we can account for this if both  $MAXSTRESS[H]$  and  $REALIZETONE$  outrank  $OBSCODA \rightarrow L$ . Associating a low tone to the coda is important, but realizing an underlying H is even more so.

(33) Contrastive high tone in CVO syllables

| /CVO-H/                                                                   | MAXSTRESS[H] | REALIZETONE | OBSCODA→L | *[H] | *[L] |
|---------------------------------------------------------------------------|--------------|-------------|-----------|------|------|
| $\begin{array}{c} \mu \mu \\ \text{CVO} \end{array}$                      | *!           |             | *         |      |      |
| $\begin{array}{c} L \\ / \backslash \\ \mu \mu \\ \text{CVO} \end{array}$ | *!           |             |           |      | *    |
| $\begin{array}{c} H L \\     \\ \mu \mu \\ \text{CVO} \end{array}$        |              | *!          |           | *    | *    |
| $\begin{array}{c} H \\ / \backslash \\ \mu \mu \\ \text{CVO} \end{array}$ |              |             | *         | *    |      |

A falling contour, with the L doubly-linked to the obstruent and to the single sonorant mora, is ruled out in CVO syllables by the high-ranked prohibition on two tones associating to a single mora,  $*[TT]\mu$ . (We saw above that we needed this constraint to rule out complex fall-rise contours on longer syllables.) In high-toned syllables with two sonorant morae,  $ObsCoda \rightarrow L$  and  $MaxStressH$  can both be satisfied, and the result is a falling contour (29 and 30). In CVO syllables, only one tone can surface, and the higher ranking of  $MaxStress[H]$  ensures that that tone will be H. Tableau (34) shows that even if the input is a contour, H will surface on CVO syllables.

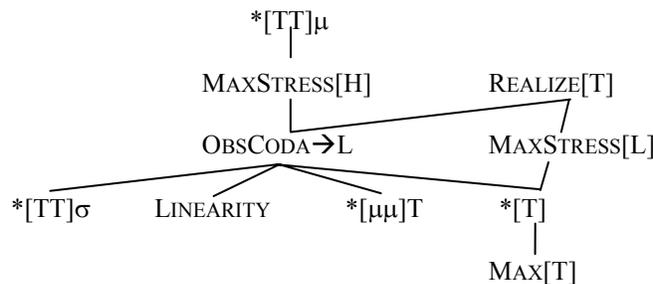
(34) Neutralization to high in CVO syllables

| /CVO-HL/ | $*[TT]\mu$ | $MAXSTRESS[H]$ | REALIZE[TONE] | $MAXSTRESS[L]$ | $OBSCODA \rightarrow L$ |
|----------|------------|----------------|---------------|----------------|-------------------------|
|          |            |                | *!            |                |                         |
|          | *!         |                |               |                |                         |
|          |            | *!             |               |                |                         |
|          |            |                |               | *              | *                       |

### 7. Conclusions

We have proposed the constraint ranking shown in (35). This ranking yields the correct result for each syllable type. In the CVV, CVVS, and CVS cases, the faithful output is always optimal. In CVVO,  $OBSCODA \rightarrow L$  requires low or falling tones. In CVO, only one tone can be realized. If there is an H in the input, high-ranking  $MAXSTRESS[H]$  ensures it will be realized. Otherwise,  $OBSCODA \rightarrow L$  provides L.

(35) Constraint ranking for Thai tones



*Lexical Tone and Markedness in Standard Thai*

In addition, we have argued for a representation where the mora is the tone-bearing unit, and mid tone is represented by the absence of phonological tone specification—points that have been disputed in the literature on Thai phonology. Finally, but not least importantly, we believe we have neatly solved a tricky distributional puzzle.

**References**

- Abramson, A. 1962. *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments*. Mouton and Co.
- Alderete, J. 1995. Faithfulness to Prosodic Heads. Ms., University of Massachusetts, Amherst.
- Beckman, J. 1995. Shona Height Harmony: Markedness and Positional Identity. In J. Beckman, L. Dickey, and S. Urbanczyk (eds.), *University of Massachusetts Occasional Papers: Papers in Optimality Theory* 18:53-76.
- Broselow, E., S. Chen, and M. Huffman. 1998. Syllable Weight: Convergence of Phonology and Phonetics. *Phonology* 14:47-82.
- Diller, A. 1996. Thai and Lao writing. In P. Daniels and W. Bright (eds.), *The World's Writing Systems*. New York: Oxford University Press.
- Erickson, D., and A. Abramson. 1972. Electromyographic Study of the Tones of Thai. *Haskins Laboratory: Status Report on Speech Research* SR-31/32.
- Gandour, J. 1974. On the Representation of Tone in Siamese. *UCLA Working Papers in Phonetics* 27:118-146.
- Kruatrachue, F. 1960. Thai and English: A Comparative Study of Phonology for Pedagogical Applications. Ph.D. diss., Indiana University.
- Maddieson, I. 1976. A Further Note on Tone and Consonants. *UCLA Working Papers in Phonetics* 33:131-159.
- Maddieson, I. 1977. Tone Effects on Consonants. *UCLA Working Papers in Phonetics* 36:91-110.
- Morén, B. 1999. Distinctiveness, Coercion and Sonority: A Unified Theory of Weight. Ph.D. diss., University of Maryland at College Park.
- Yip, M. 1982. Against a Segmental Analysis of Zhaoh and Thai: A Laryngeal Tier Proposal. *Linguistic Analysis* 9:79-94.

Center for Advanced Study in  
Theoretical Linguistics  
University of Tromsø  
9037 Tromsø  
Norway

bruce.moren@hum.uit.no

Department of Linguistics  
Box 571051  
Georgetown University  
Washington, DC 20057  
USA

zsigae@georgetown.edu



# The Relationship between the Source Characteristic and Frequency

HANSANG PARK  
*University of Texas at Austin*

## 0. Introduction

The Source Filter Theory assumes that the amplitude level of the source characteristics of the harmonics falls off at the approximate rate of -12 dB/octave in the source spectrum.

The combined source and radiation characteristics constitute a spectrum that falls off at the approximate rate of 6 dB/octave.  $|R(f)|$  is proportional to frequency,  $f$ , and it will be assumed that  $|U(f)|$  is approximately proportional to  $1/f^2$  above cutoff frequency of 100 c/s. This relation can thus be written

- (1) The combined source and radiation characteristics

$$|U(f)||R(f)| = P_k \frac{(f/100)}{1 + (f/100)^2}, \quad (1.3-2)$$

where brackets indicate absolute values and  $(f)$  function of frequency,  $|U(f)|$  symbolizes the amplitude-versus-frequency characteristics of the source,  $|R(f)|$  the frequency characteristic of radiation, and  $P_k$  is a constant determining the particular sound pressure level. (Fant 1960:49)

This paper will make it clear that the source rate is not fixed to -12 dB/octave but varies with frequency, such that the source rate is not a constant but a variable. Specifically, this paper will first derive the ratio between the source characteristics of the first two harmonics.<sup>1</sup> Second, this paper will calculate the rate in dB/octave from the ratio between the source characteristics of the first two harmonics at certain frequencies and illustrate the relationship between the source rate and frequency. Finally, this paper will reconsider the existing metrics for phonation type that employ the first two harmonics of the source spectrum and, furthermore, propose a new metric for phonation type: *Phonation Type Index K*.

---

<sup>1</sup> The discussion of the ratio will be restricted to the first two harmonics for the time being, so that only  $f_0$  and  $2f_0$  will appear in the discussion.

The finding that the source rate is not fixed to -12 dB/octave but varies with frequency is important in two respects. First, the ratio and the rate can be applied to the whole range of frequencies. It is not necessary to fix a cutoff frequency. The only limitation is the range of frequency within which human beings can produce speech sounds (humans are generally not thought to produce speech sounds with a fundamental frequency below 50 Hz or above 2000 Hz).

Second, it is necessary to reconsider the measures of phonation type, such as H1-H2, H1\*-H2\*, and Cor(H1-H2), considering the contribution of the varying rate to the amplitude of the harmonics. The difference in amplitude between the first two harmonics has been used as a metric of phonation type. H1-H2 was an acceptable metric of the phonation type for low vowels minimally affected by the first formant, while it was not appropriate for high vowels, since F1 boosts the amplitude level of the first two harmonics. H1\*-H2\* was a corrected metric of the phonation types (Stevens and Hanson 1995), but details of the metric are not available. Cor(H1-H2) was recently designed by Ahn (1999) based on the Source Filter Theory. It has two advantages over H1-H2. One is that the contributions of filter and radiation characteristics can be removed. The other is that it is possible to compare different phonation types in terms of the relative difference between the observed H1-H2 and the expected H1-H2, even though the absolute values of the source characteristics of the harmonics are not available. However, these measures did not clarify the contribution of the varying rate that may lead to a significantly different result. If we take the varying rate into consideration, we can provide more reliable data with which we can determine the significance of the difference.

## **1. The Source Characteristic and Frequency**

### **1.1. The Ratio between $|U(f)|$ and $|U(2f)|$**

This section shows how the ratio between the source characteristics of the first two harmonics was derived. The derivation of the ratio between  $|U(f)|$  and  $|U(2f)|$  is shown in (2). It should be noted that the ratio between the source characteristics of the first two harmonics was derived from the formula in (1).

The first two rows in (2-1) and (2-2) represent the combined source and radiation characteristics at the fundamental frequency  $f_0$  and at double the fundamental frequency  $2f_0$ . The second row is obtained by substituting  $f_0$  with  $2f_0$ .

*The Source Characteristic and Frequency*

(2) Derivation of the ratio between  $|U(f)|$  and  $|U(2f)|$

$$(2-1): |U(f_0)||R(f_0)| = P_k \frac{100(f_0)}{10,000 + (f_0)^2}$$

$$(2-2): |U(2f_0)||R(2f_0)| = P_k \frac{200(f_0)}{10,000 + 4(f_0)^2}$$

$$(2-3): \frac{|U(2f_0)||2R(f_0)|}{|U(f_0)||R(f_0)|} = \frac{P_k \frac{200(f_0)}{10,000 + 4(f_0)^2}}{P_k \frac{100(f_0)}{10,000 + (f_0)^2}}$$

$$(2-4): \frac{|U(2f_0)||2R(f_0)|}{|U(f_0)||R(f_0)|} = \frac{\frac{200(f_0)}{10,000 + 4(f_0)^2}}{\frac{100(f_0)}{10,000 + (f_0)^2}}$$

$$(2-5): \frac{|U(2f_0)||R(2f_0)|}{|U(f_0)||R(f_0)|} = \frac{\frac{2}{10,000 + 4(f_0)^2}}{\frac{1}{10,000 + (f_0)^2}}$$

$$(2-6): \frac{|U(2f_0)||R(2f_0)|}{|U(f_0)||R(f_0)|} = \frac{2(10,000 + (f_0)^2)}{10,000 + 4(f_0)^2}$$

$$(2-7): \frac{|U(f_0)|}{|U(2f_0)|} \frac{2}{1} = \frac{2(10,000 + (f_0)^2)}{10,000 + 4(f_0)^2}$$

$$(2-8): \frac{|U(2f_0)|}{|U(f_0)|} = \frac{10,000 + (f_0)^2}{10,000 + 4(f_0)^2}$$

The third row in (2-3) represents the ratio between the combined source and radiation characteristics at  $f_0$  and at  $2f_0$ . The third row in (2-3) is obtained by dividing the combined source and radiation characteristics at  $2f_0$  by those at the fundamental frequency  $f_0$ . The fourth and fifth rows in (2-4) and (2-5) show the

arrangement of the right-hand side by canceling  $P_k$  and  $100(f_0)$ , which both the numerator and denominator have in common. The row in (2-6) is the result of the computation. It is necessary to remove the ratio between  $|R(2f_0)|$  and  $|R(f_0)|$ , since we are interested in the ratio between  $|U(2f_0)|$  and  $|U(f_0)|$ . The ratio between  $|R(2f_0)|$  and  $|R(f_0)|$  is 2. The row in (2-7) represents the substitution of the ratio of the radiation characteristics by 2. The row in (2-8) shows the ratio between the source characteristics of the first two harmonics obtained by canceling the 2 in both sides. It is apparent from the derivation in (2) that the ratio is a function of frequency.

### 1.2. The Source Rate

It is necessary to examine the source rates to see how the source rate varies with frequency. The source rate is calculated by taking the logarithm with base 10 of the ratio between the source characteristics of the first two harmonics and multiplying 20 to the logarithm. For example, the source rate in dB/octave equals -7.96, as seen in (3).

(3) The source rate in dB/octave at 100 Hz

$$\frac{|U(200)|}{|U(100)|} = \frac{10,000 + 100^2}{10,000 + 4(100)^2} = \frac{2}{5} = .4$$

$$20 \log_{10} \frac{|U(200)|}{|U(100)|} = -7.96$$

As another example, the source rate when the frequency approaches infinity is -12.04 dB/octave, as seen in (4).

(4) The source rate in dB/octave when frequency approaches infinity

$$\lim_{f \rightarrow \infty} \frac{|U(2f)|}{|U(f)|} = \lim_{f \rightarrow \infty} \frac{(10,000 + f^2)}{(10,000 + 4f^2)} = \frac{1}{4} = .25$$

$$20 \log_{10} \lim_{f \rightarrow \infty} \frac{|U(2f)|}{|U(f)|} = 20 \log_{10} \lim_{f \rightarrow \infty} \frac{(10,000 + f^2)}{(10,000 + 4f^2)}$$

$$= -12.04$$

In the same way, we can obtain the source rates at given frequencies. The source rates in dB/octave at some frequency values are given in (5).

*The Source Characteristic and Frequency*

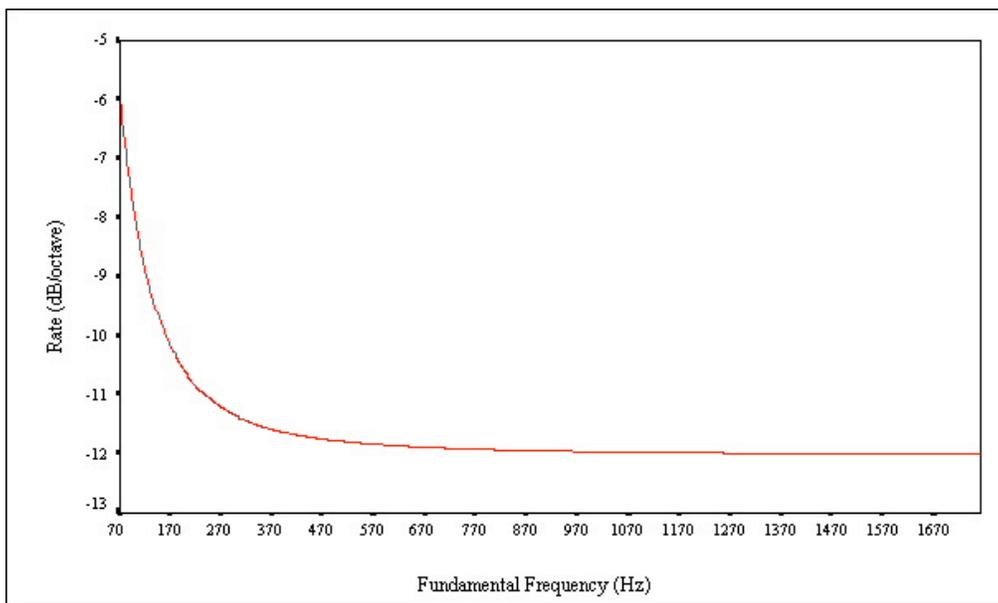
(5) The source rates in dB/octave at given frequencies

| Frequency in Hz | Ratio | Rate in dB/octave | Remarks                        |
|-----------------|-------|-------------------|--------------------------------|
| 1               | 1.00  | 0                 | Lowest positive integer        |
| 20              | 0.90  | -0.95             | Lower threshold of audibility  |
| 70              | 0.63  | -5.96             | Lowest F0 of human voice       |
| 100             | 0.40  | -7.96             |                                |
| 150             | 0.33  | -9.79             | Mean of male voice             |
| 225             | 0.29  | -10.89            | Mean of female voice           |
| 250             | 0.28  | -11.09            | Mean of children's voice       |
| 20000           | 0.25  | -12.04            | Higher threshold of audibility |
| $\infty$        | 0.25  | -12.04            | Infinity                       |

As seen in (5), the source rate in dB/octave at mean fundamental frequencies of male, female, and children's voices, which are estimated to occur around 150, 225, and 250 Hz, respectively, are -9.79 for male voices, -10.89 for female voices, and -11.09 for children's voices, respectively. On the other hand, the source rates in dB/octave at given extreme values of frequencies are also available. The source rates in dB/octave at the thresholds of audibility, that is 20 Hz and 20000 Hz, are -0.95 and -12.04, respectively. The source rate in dB/octave at the lowest integer, or in the case where the frequency equals 1, is 0.

It is apparent that the rate is not fixed to -12 dB/octave but varies with frequency; the rate is not a constant but a variable. The relationship between the source rate and frequency is illustrated in (6).

(6) The source rate in dB/octave over fundamental frequency



In (6), the x-axis represents frequency, while the y-axis represents the source rate in dB/octave. The range of frequency is 1700 Hz, since the minimum fundamental frequency human beings can produce is estimated to be 70 Hz and the maximum fundamental frequency is estimated to be 1770 Hz. It should be noted that ordinary speech sounds seldom exceed 1000 Hz. We can see that the source rate varies dramatically in lower frequencies. The value of -12 dB/octave is never applicable to lower frequencies below 200 Hz, rather only to frequencies higher than 500. The fact that male voices seldom reach that frequency suggests that male voices are seriously influenced by the varying source rate. The lower the frequency, the more significant the effect of the rate is.

## 2. Reconsideration of the Metrics of Phonation Type

Since the source rate is not fixed to -12 dB/octave but varies with frequency, the source rate is not a constant but a variable. This fact leads us to reconsider the metrics of phonation type, since the single most widely used metric has been difference in amplitude between the first two harmonics—that is, H1-H2—and the varying source rate may significantly affect the values of H1-H2. In this section, such metrics of phonation type<sup>2</sup> as H1-H2 (Huffman 1987) and Cor(H1-H2) will be examined.

### 2.1. H1-H2

Phonation type has been measured in terms of difference in amplitude between the first two harmonics—that is, H1-H2 (Huffman 1987). H1 and H2 can be represented as in (7), respectively.

(7) Representation of H1 and H2

$$HARMONIC_1 = H_1 = 20 \log_{10} |Q(f_0)U(f_0)||R(f_0)||H_n(f_0)|$$

$$HARMONIC_2 = H_2 = 20 \log_{10} |Q(2f_0)U(2f_0)||R(2f_0)||H_n(2f_0)|$$

This representation is based on the Source Filter Theory where sound pressure is the product of source, radiation, and filter characteristics. Terms for phonation type,  $Q(f_0)$  and  $Q(2f_0)$ , were inserted into the source characteristic, since phonation type was basically considered to be a factor of the source characteristic.

H1-H2 can be calculated by subtracting H2 from H1. H1-H2 is represented in (8).

---

<sup>2</sup> Stevens and Hanson (1995) also proposed H1\*-H2\*. However, details of the metric are not available.

(8) Representation of H1-H2

$$H_1 - H_2 = A + B + C + D$$

$$A = DIFF(PHONATIONTYPE) = 20 \log_{10} |Q(f_0)| - 20 \log_{10} |Q(2f_0)|$$

$$B = DIFF(SOURCERATE) = 20 \log_{10} |U(f_0)| - 20 \log_{10} |U(2f_0)|$$

$$C = DIFF(RADIATION) = 20 \log_{10} |R(f_0)| - 20 \log_{10} |R(2f_0)| = 6dB/octave$$

$$D = DIFF(TRANSFER) = 20 \log_{10} |H_n(f_0)| - 20 \log_{10} |H_n(2f_0)|$$

As seen in (8), H1-H2 can be represented as the sum of four differences: differences in phonation type (A), the source rate (B), the radiation characteristic (C), and vocal tract transfer function (D). The difference in the radiation characteristic is fixed at 6 dB/octave. The source rate (B) varies with F0 and the vocal tract transfer function (D) also varies with F0 as well as formant frequencies, which result from a difference in vowel quality. However, H1-H2 did not remove the contributions of the radiation characteristic, the vocal tract transfer function, or the varying source rate. H1-H2 is acceptable only if B and D stay the same across the speech samples.

**2.2. Cor(H1-H2)**

Phonation type has also been measured in terms of relative difference between the observed difference between the first two harmonics (Obs(H1-H2)) and the expected difference between the first two harmonics (Exp(H1-H2)) (Ahn 1999). Exp(H1-H2) was subtracted from Obs(H1-H2) to remove the contributions from the radiation and the vocal tract transfer function. Observed H1 and H2 can be represented as in (9).

(9) Representation of Obs(H1) and Obs(H2)

$$Obs(H_1) = 20 \log_{10} \left| Q(f_0) P_k \frac{(f_0/100)}{1 + (f_0/100)^2} \right| |R(f_0)| |H(f_0)|$$

$$Obs(H_2) = 20 \log_{10} \left| Q(2f_0) P_k \frac{(f_0/100)}{1 + (2f_0/100)^2} \right| |R(2f_0)| |H(2f_0)|$$

Obs(H1) and Obs(H2) are the same as H1 and H2 in (9), respectively, except that the source characteristics are divided into the constant determining the particular sound pressure level ( $P_k$ ) and frequency response of the source characteristic. On the other hand, Exp(H1) and Exp(H2) are represented in (10).

(10) Representation of Exp(H1) and Exp(H2)

$$Exp(H_1) = 20 \log_{10} \left| Q'(f_0) P_k \frac{(f_0/100)}{1 + (f_0/100)^2} \right| |R(f_0)| |H(f_0)|$$

$$Exp(H_2) = 20 \log_{10} \left| Q'(2f_0) P_k \frac{(f_0/100)}{1 + (2f_0/100)^2} \right| |R(2f_0)| |H(2f_0)|$$

Ahn (1999) assumed that the vocal tract transfer functions of Exp(H1) and Exp(H2) are the same as those of Obs(H1) and Obs(H2). He also assumed that difference in phonation type equals 0—that is, that  $Q'(f_0)$  equals  $Q'(2f_0)$ . In addition, he fixed the source rate to -12 dB/octave and set  $P_k$  to 1.

Based on the representations given in (9) and (10), Obs(H1-H2) and Exp(H1-H2) can be represented as in (11) and (12), respectively.

(11) Obs(H1-H2)

$$Obs(H_1 - H_2) = A + B + C + D$$

$$A = DIFF(PHONATIONTYPE) = 20 \log_{10} |Q(f_0)| - 20 \log_{10} |Q(2f_0)|$$

$$B = b1 + b2$$

$$b1 = DIFF(PRESSURE) = 20 \log_{10} P_k - 20 \log_{10} P_k$$

$$b2 = DIFF(SOURCERATE) = -20 \log_{10} \left| \frac{1 + (f_0/100)^2}{1 + (2f_0/100)^2} \right|$$

$$C = DIFF(RADIATION) = 20 \log_{10} |R(f_0)| - 20 \log_{10} |R(2f_0)| = 6dB / octave$$

$$D = DIFF(TRANSFER) = 20 \log_{10} |H(f_0)| - 20 \log_{10} |H(2f_0)|$$

(12) Exp(H1-H2)

$$Exp(H_1 - H_2) = A + B + C + D$$

$$A = DIFF(PHONATIONTYPE) = 20 \log_{10} |Q'(f_0)| - 20 \log_{10} |Q'(2f_0)| = 0$$

$$B = b1 + b2$$

$$b1 = DIFF(PRESSURE) = 20 \log_{10} 1 - 20 \log_{10} 1 = 0$$

$$b2 = DIFF(SOURCERATE) = -20 \log_{10} \left| \frac{1 + (f_0/100)^2}{1 + (2f_0/100)^2} \right| = -12dB / octave$$

$$C = DIFF(RADIATION) = 20 \log_{10} |R(f_0)| - 20 \log_{10} |R(2f_0)| = 6dB / octave$$

$$D = DIFF(TRANSFER) = 20 \log_{10} |H(f_0)| - 20 \log_{10} |H(2f_0)|$$

### *The Source Characteristic and Frequency*

As seen in  $b1$  in (12), it is unnecessary to set  $P_k$  to 1, since  $P_k$  is cancelled by subtracting  $H2$  from  $H1$ . It is improper to set  $P_k$  to 1, since that means that there is no audible pressure. It is also improper to set the source rate to -12 dB/octave, since it varies with frequency.

On the other hand,  $Cor(H1-H2)$  can be obtained by subtracting  $Exp(H1-H2)$  from  $Obs(H1-H2)$ , which is shown in (13).

(13)  $Cor(H1-H2)$

$$\begin{aligned}
 & Cor(H1 - H2) \\
 &= Obs(H_1 - H_2) - Exp(H_1 - H_2) \\
 &= DIF(A) + DIF(B) + DIF(C) + DIF(D) \\
 \\ 
 & DIF(A) = (20 \log_{10} |Q(f_0)| - 20 \log_{10} |Q(2f_0)|) - 0 \\
 &= 20 \log_{10} |Q(f_0)| - 20 \log_{10} |Q(2f_0)| \\
 & DIF(B) = DIF(b1) + DIF(b2) \\
 & DIF(b1) = 0 - 0 = 0 \\
 & DIF(b2) = -20 \log_{10} \left| \frac{1 + (f_0 / 100)^2}{1 + (2f_0 / 100)^2} \right| + 12 \\
 & DIF(C) = 6 - 6 = 0 \\
 & DIF(D) = 0
 \end{aligned}$$

As seen in (13),  $Cor(H1-H2)$  is the sum of the differences in phonation type and the source rate between  $Obs(H1-H2)$  and  $Exp(H1-H2)$ . The differences in phonation type and the source rate vary with fundamental frequency, while all other differences equal 0. It was claimed that  $Exp(H1-H2)$  was subtracted from  $Obs(H1-H2)$  to remove the contribution of the vocal tract transfer function. However, it seems that the vocal tract transfer function in  $Obs(H1-H2)$  was substituted with the calculated values of the vocal tract transfer function in  $Exp(H1-H2)$ . It was stated that the source rate was fixed to -12 dB/octave, but in reality, the source rate was not fixed to -12 dB/octave, since  $Cor(H1-H2)$  was calculated by subtracting  $Exp(H1-H2)$  from  $Obs(H1-H2)$  without fixing the source rate to -12 dB/octave. Ahn (1999) seems to have been unaware of the varying source rate. However,  $Cor(H1-H2)$  is well designed in the sense that it tried to remove the contribution of the filter characteristic.  $Cor(H1-H2)$  can be a good metric of phonation type for speech samples with varying  $F0$  and formant frequencies only if  $Cor(H1-H2)$  incorporates the varying source rate into the metric.

### 3. Phonation Type Index $K$

It was noted that  $\text{Cor}(H1-H2)$  could serve as a metric of phonation type if it incorporates the varying source rate into the metric without fixing the source rate to -12 dB/octave. However, it only describes how speech samples are different with respect to phonation type. It does not explain why spectral tilt varies with phonation types. In this section, a general term for harmonics will be presented to explain why phonation type varies across speech samples. A general term for harmonics is given in (14).

(14) A general term for harmonics

$$HARMONIC_n(f_0) = SOURCE(f_0) + RADIATION + TRANSFER(f_0)$$

$$SOURCE(f_0) = 20 \log_{10} \left| P_k \frac{(f_0/100)}{1+(f_0/100)^2} \frac{1+(f_0/100)^2}{1+(nf_0/100)^2} \left(\frac{1}{n}\right)^K \right|$$

$$RADIATION = 20 \log_{10} n$$

$$TRANSFER(f_0) = 20 \log_{10} |H(nf_0)|$$

As seen in (14), the source characteristic is the product of the constant determining the particular sound pressure level, the frequency response at the fundamental frequency, the source rate, and the phonation type. It should be noted that the combined source and radiation characteristics was first divided into the radiation characteristic and the source characteristic which was, in turn, divided into the frequency response of the fundamental component and the source rate. The only difference between the present model and the Source Filter Theory is the term for phonation type, that is,  $(1/n)^K$ . This term was established under the rationale that it is responsible for spectral tilt and, therefore, varies with phonation type, and that it contributes nothing to the spectral tilt in H1 or when  $K$  equals 0. According to the general term for harmonics, H1 and H2 can be represented as in (15).

(15) Representation of H1 and H2

$$H_1(f_0) = 20 \log_{10} \left| P_k \frac{(f_0/100)}{1+(f_0/100)^2} \right| |H(f_0)|$$

$$H_2(f_0) = 20 \log_{10} \left| P_k \frac{(f_0/100)}{1+(f_0/100)^2} \frac{1+(f_0/100)^2}{1+(2f_0/100)^2} \left(\frac{1}{2}\right)^K \right| 2 |H(2f_0)|$$

Difference in phonation type can be calculated by subtracting H2 from H1. The difference in phonation type is represented in (16).

(16) Difference in phonation type

$$[H_1 - H_2](f_0) = A + B + C + D$$

$$A = \text{DIFF}(\text{PHONATIONTYPE}) = -20 \log_{10} \left(\frac{1}{2}\right)^K$$

$$B = \text{DIFF}(\text{SOURCERATE}) = -20 \log_{10} \frac{10,000 + (f_0)^2}{10,000 + 4(f_0)^2}$$

$$C = \text{DIFF}(\text{RADIATION}) = -20 \log_{10} 2$$

$$D = \text{DIFF}(\text{TRANSFER}) = 20 \log_{10} |H(f_0)| - 20 \log_{10} |H(2f_0)|$$

$$\text{DPT}(f_0) = A = [H1 - H2](f_0) - (B + C + D)$$

$$\text{PTI } K = \frac{A}{6.02}$$

Difference in phonation type (DPT) can be obtained by subtracting the difference in the source rate (B), the radiation characteristic (C), and the filter characteristic (D) from the measured values of H1-H2. Furthermore, phonation type index  $K$  can be obtained by dividing the difference in phonation type (A) by 6.02. DPT is the same as Cor(H1-H2) except that phonation type systematically controls the source rate and difference in phonation type is represented by phonation type index  $K$ . It should be noted that fundamental frequency must always be specified, since the source rate can be calculated only when fundamental frequency is specified. It is important to take the varying source rate into account when phonation types are compared across speech samples with substantially different fundamental frequency. Taking the varying source rate into consideration, we can provide more reliable data by which the significance of difference in phonation type can be determined.

#### **4. Conclusion**

The ratio between the source characteristics of H1 and H2 was derived. It was shown that the source rate varies with frequency. This finding was significant, since the varying source rate affects the amplitude difference between the first two harmonics. Metrics of phonation type, such as H1-H2 and Cor(H1-H2), were evaluated from this perspective. It was noted that it is important to incorporate the varying source rate into the metric of phonation type. In addition, special attention needs to be paid when speech samples with a substantial difference in fundamental frequency are compared. The existing metrics of phonation type do not account for why spectral tilt varies with phonation type. A new metric of phonation type, phonation type index  $K$ , was designed to account for the difference in spectral tilt. It has advantages over H1-H2 or Cor(H1-H2) in that it can account for variation in spectral tilt not only at the first two harmonics but also at higher

Hansang Park

harmonics. However, much empirical data are necessary to validate DPT and phonation type index *K*. This remains to be done.

### References

- Ahn, Hyunkee. 1999. Post-Release Phonatory Processes in English and Korean: Acoustic Correlates and Implications for Korean Phonology. Ph.D. diss., University of Texas at Austin.
- Fant, Gunnar. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Huffman, Marie. 1987. Measures of Phonation Type in Hmong. *Journal of the Acoustical Society of America* 81:495-504.
- Stevens, Kenneth N., and Helen Hanson. 1995. Classification of Glottal Vibration from Acoustic Measurements. In O. Fujimura and M. Hirano (eds.), *Vocal Fold Physiology*. San Diego: Singular Publishing Group, Inc.

Department of Linguistics  
Calhoun Hall 501  
The University of Texas at Austin  
Austin, TX 78712-1196

phans@mail.utexas.edu

# A Unified Account of Essentially Contested Concepts

JASON D. PATENT

*University of California, Berkeley and Stanford University*

## 0. Introduction

Cognitive linguistics has taken as one of its main challenges the explanation of category structure. By and large, the discipline has been up to the task. But a particular kind of category has eluded optimally clear explication: what W.B. Gallie (1956) originally called “essentially contested concepts”, concepts such as democracy and art, which by nature invite disagreement over their meanings. For our purposes, the topic is best referred to as “essentially contested categories”, since cognitive linguistics as a whole has a more well-problematized notion of *category* than of *concept*. For primarily stylistic reasons I will normally refer simply to *contested categories*.

Gallie’s work aroused some interest in philosophical circles, but was largely ignored by social scientists until William E. Connolly, a political scientist, applied some of Gallie’s notions to categories in his discipline (1993/1974). Eventually, contested categories caught the eye of George Lakoff, who in some unpublished remarks gave a first pass at a cognitive-linguistic analysis of their structure. These unpublished remarks were then cited by two other linguists—Alan Schwartz (1992) and Pamela Morgan (1992, 1998)—and some misconceptions from Lakoff’s original work on the topic have been perpetuated as a result. The goal of this paper is to correct these misconceptions, ultimately yielding, I hope, an analysis that is more rigorous than previous analyses in the application of the tools of cognitive linguistics.

## 1. Theoretical Background and Terminology

One of George Lakoff’s contributions to the world of linguistics has been his notion of *radial category*. Radial categories received their most thorough explication in Lakoff (1987). The relatively simple example he gives of a radial category is *mother* (Lakoff 1987:74-76). The overall idea is this: there are several frames, or submodels, that constitute the category. Given a person, say Chris, the mother of Chris could be the person who contributed half of Chris’s genetic material; gave birth to Chris; is the primary female nurturer of Chris; or is married to Chris’s father. There are thus four submodels: genetic, birth, nurturance, and

marital. The *prototype* of the category is where all the submodels converge. That is, in the prototypical case, Chris's mother gave birth to him, provided half of his genes, is Chris's primary female nurturer, and is married to Chris's father. Non-prototypical, or peripheral, cases involve only some of the submodels. For example, Chris's stepmother only adheres to the marital model (though nurturance is probably involved as well), whereas his birth mother adheres to the birth model, probably the genetic model, and possibly provides some nurturance, but the marital model doesn't hold.

Crucially for our purposes, there is agreement at the center and contestedness at the periphery: no one would disagree that a woman adhering to all four submodels is Chris's "real mother," whereas there is plenty of room for disagreement about who the "real mother" is if we have a choice, say, among surrogate mother, birth mother, stepmother and adoptive mother.

This category-structure principle of convergence-plus-agreement at the center and reduction-plus-contestedness at the periphery is a useful one to bear in mind throughout the paper.

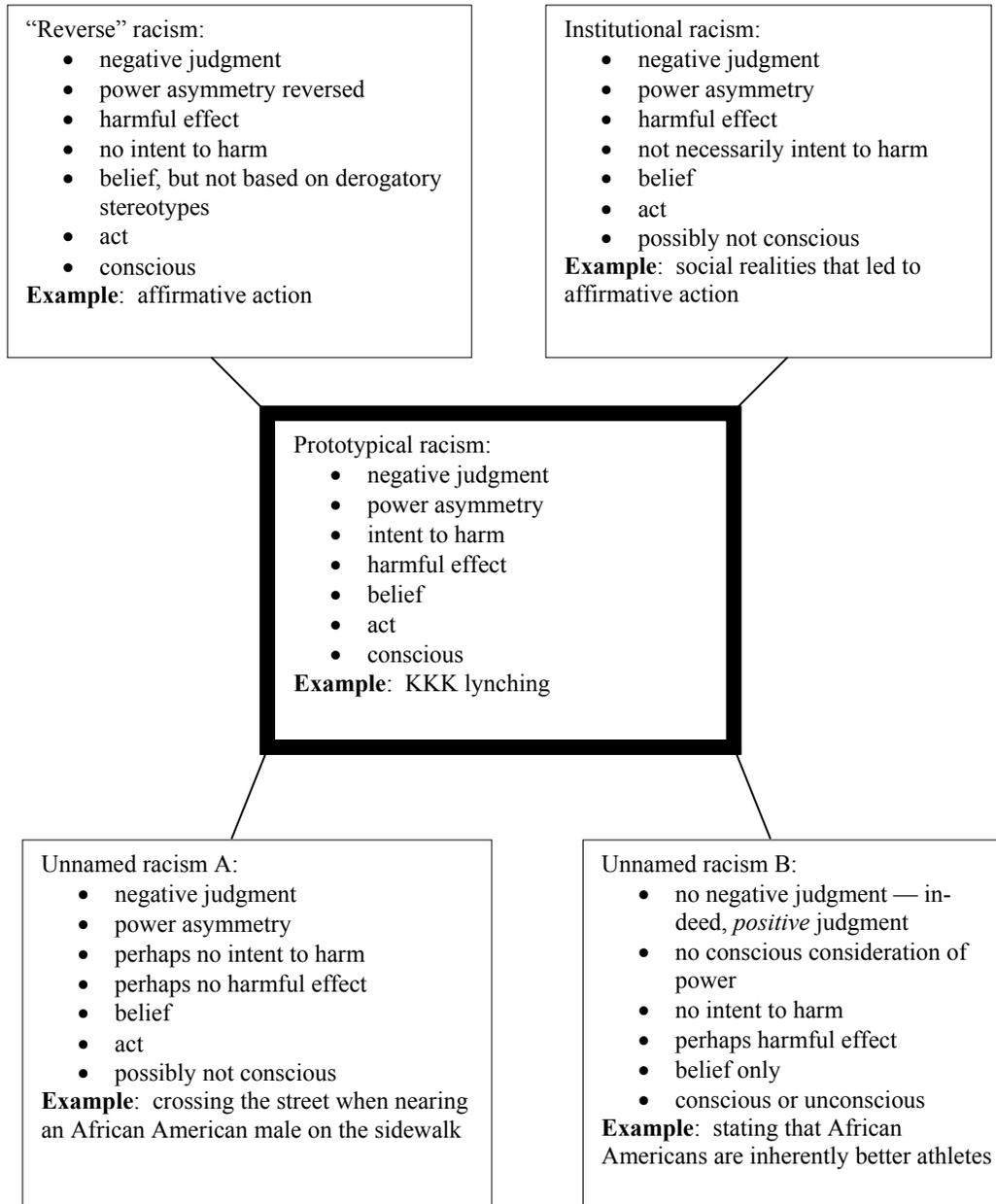
## **2. *Racism: A Case Study***

Not all of Gallie's original criteria for determining whether something is a contested category are relevant for this paper, but three certainly are. What follows is a brief description of the sort of category Gallie had in mind, focusing on these three criteria. (Here, I substitute cognitive-linguistic terminology for Gallie's original wording.) A contested category is one in which value judgments inhere; which is composed of a complex matrix of submodels; and in disagreements over which people's differences stem from the differential weighting of the submodels.

By these criteria, *racism* is a good example of a contested category. Labeling someone or something as "racist" certainly implies a value judgment, and as we will see in a moment the category is composed of a rich set of submodels that are differentially weighted for different types of racism—not all of which would be agreed upon as instances of the category.

Figure 1 below is intended to represent the structure of *racism* from this perspective. Note that in Figure 1, the submodels constituting each subtype—bulleted in each box—are assessed from the viewpoint of the person using the label *racism*, rather than from the viewpoint of the person holding the racist belief and/or committing the racist act.

**Figure 1.** Racism as a radial category



First, a caveat: this is severely oversimplified. Hopefully, though, there is enough here to make the point I want to make about contested categories and how linguists should go about analyzing them.

Let's start by looking at prototypical racism at the center. An example of this form of racism is a lynching by the Ku Klux Klan. It adheres to (at least these) seven submodels (referring here to the person whose racism is under consideration as R):

1. It involves a negative judgment by R of a person or group of people based on racial category membership.<sup>1</sup>
2. There is a power asymmetry in which R belongs to a societal group historically more enfranchised than the person against whom the act is being perpetrated.
3. R intends to cause harm.
4. R causes harm.
5. R holds certain derogatory, stereotype-based beliefs about members of a racial group.
6. R acts on these beliefs.
7. These beliefs are conscious.

The phenomenon described here would be largely agreed upon as an instance of *racism*. From the perspective of racial-category analysis, this results from the overlap of so many of the submodels that are relevant to the category *racism*.

But what of cases where such overlap isn't present? We can look to the peripheral examples in Figure 1. For instance, institutional racism, in the upper right-hand box, involves either five or six of the submodels: it is conceivable that someone would argue that institutional racism isn't racism because it doesn't involve intent to harm. To take another example, affirmative action—which some claim is a form of “reverse racism”—adheres to only four of the seven submodels. From the standpoint of someone labeling it as “racist,” affirmative action: (i) involves negative judgment of a person based on racial category membership (“negative” in the sense of “less qualified” by whatever relevant criteria), (ii) has a harmful effect on members of a racial category, (iii) is a set of acts carried out (as opposed to simply a belief), and (iv) is consciously applied. But three submodels are missing or crucially modified. First, the original power asymmetry is reversed; second, there is no intent to harm anyone because of his or her membership in a racial category;<sup>2</sup> and third, while based on “belief” of some sort, “reverse racism” isn't based on the same sort of derogatory racial category-based stereotyping of prototypical racism.

---

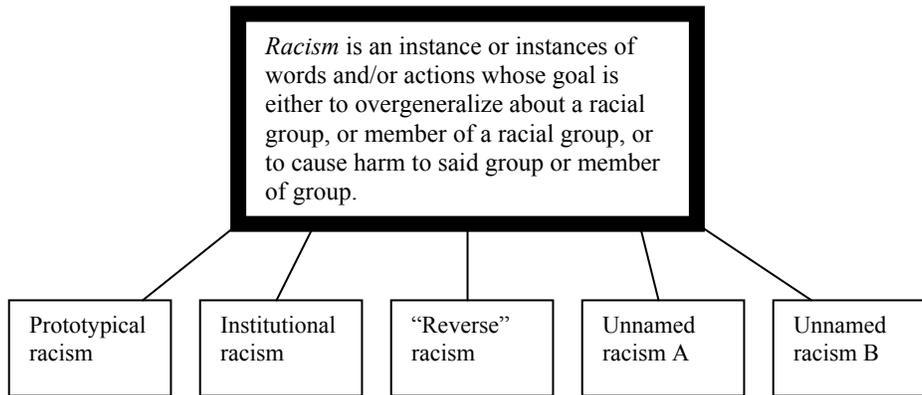
<sup>1</sup> How racial category membership is determined is itself an enormously complex issue deserving of separate treatment. I have taken a preliminary look at some of the relevant issues in Patent (1997).

<sup>2</sup> Though it is possible to imagine diehard affirmative action opponents disagreeing.

While many of the details of *how* this category is structured are debatable, the overall point is to show *that* radial categories can be represented this way: sub-model overlap and agreement maximized at the center and lessened at the periphery.

Another way to represent the structure of *racism* is given below, in Figure 2.

**Figure 2.** Racism as schema and subtype



In this Langackerian (1987) representation, subtypes of racism are categorized as subordinate to a schematic representation that generalizes over all the subcases.

What we will see below is that any contested category can be represented either as a radial category, as in Figure 1, or as a schema, as in Figure 2. This allows us to eliminate a troubling dichotomy that has persisted for years in the study of contested categories within cognitive linguistics.

### **3. Problems with Previous Analyses**

In some unpublished remarks,<sup>3</sup> George Lakoff proposed a distinction between two types of contested categories (referred to in his remarks as “contested concepts”). Of the first, Type 1, he said:

There is an underspecified central model that is generally shared. Different belief systems give rise to extensions of the central model. The extensions are the contested versions of the concept...the concept is defined with respect to a relatively underspecified cognitive model—that is, a model that imposes relatively few constraints on the concept.

Of Type 2, Lakoff said that there is:

---

<sup>3</sup> In citing unpublished remarks, I am going against standard academic practice, and in doing so am being unfair to George Lakoff. I do this only because these remarks have in turn been cited by Schwartz and Morgan in their work on contested categories, and have thus become part of the literature that needs to be addressed in any study of contested categories.

...a very rich central model defined by a large cluster of simple models. The contested versions of the central model are extensions defined by one or more members of the cluster. The concept is contested because different people believe that different members of the cluster “correctly” define the concept.

Something inherently unsatisfying about this account is that, if we take the term *central* to mean the same thing for both types of contested categories, we seem to be faced with two fundamentally different sorts of categories: one with an under-specified center, whose peripheral members add specifications, and one with a rich center, whose peripheral members sift away specifications. This is troubling: why should there be two radically different *kinds* of categories, especially when one of them (Type 1) flies in the face of what we know of radial category structures, namely that the center is where the structure is richest?

To help sort this out, I now introduce a theoretical distinction from cognitive anthropology: *intrapersonal* versus *extrapersonal*. The terms come from Strauss and Quinn (1997), but these ideas under different names also figure prominently in Shore (1996). *Intrapersonal* refers to the cognitive content of an individual; *extrapersonal* refers to what is shared among individuals in a cultural community.<sup>4</sup> These are two fundamentally different perspectives on category structure, and need to be kept distinct.

We can look at both Figure 1 and Figure 2 from either perspective. From an intrapersonal standpoint, Figure 1 focuses on the submodels that converge to structure a prototype *in the mind of an individual*, and how lack of convergence of submodels yields peripheral category members, members that would likely get lower category ratings in Rosch-style (1975) prototype experiments. Figure 2, on the other hand, emphasizes, as Langacker (1987) has pointed out, that we can store knowledge at different levels of schematicity.

Turning to an extrapersonal perspective, Figures 1 and 2 mean something different. The prototype at the center of Figure 1 represents the version least likely to be contested, and most likely to be shared—contrarywise for the extensions. Figure 2 represents what is shared across all versions of *racism* held by anyone.

Equipped now with the intrapersonal/extrapersonal distinction, we can make sense of Lakoff’s positing two types of contested categories. A contested category looks like it belongs to Type 1 when seen from the extrapersonal perspective of generalizing over the category structures of many individuals. This inevitably leads to underspecification, yielding the Langackerian schema structure of Figure 2. A contested category looks like it belongs to Type 2 when seen from an intrapersonal radial-category perspective: submodels converge at the center to form an individual’s prototype, and are winnowed away at the periphery to form non-prototypical category members.

---

<sup>4</sup> This is actually not quite what is meant by *extrapersonal*. The term more properly refers to the dialectical counterpart of the intrapersonal—the publicly-available cultural forms that help construct, and are reconstructed by, the intrapersonal. Sharedness is thus but one aspect of the extrapersonal.

*A Unified Account of Essentially Contested Concepts*

If we take a closer look at Lakoff's original remarks, we can see some confusion revealed in a lack of careful attention to terminology. First, Lakoff uses *central* in two very different ways: once meaning *schema* and once meaning *prototype*. Second, he mixes together terms from the ideally separate realms of the intrapersonal and the extrapersonal. In his discussion of Type 1, for instance, he writes of "different belief systems" and "contested versions"—squarely belonging with the extrapersonal—but then mentions an "underspecified cognitive model", which seems clearly to refer to the intrapersonal.

Unfortunately, this confusion persists in Alan Schwartz's (1992) study of contested categories. Schwartz (1992:26) writes: "A contested concept is a radial category which is generated by a central ICM which is subject to contention. The central model is extended in a number of possible ways, and these fully instantiated extensions are the versions of the concepts which conflict." This seems to be about the extrapersonal, but, if so, do we want to use the term *radial category* in this way? And what exactly does Schwartz mean by "central" here? Later, Schwartz writes: "At the core of the concept of feminism, then, is an underspecified propositional model" (55) and "These questions define the shape of the slots left unfilled in the underspecified model which stands at the core of the concept of feminism" (56). As with Lakoff, Schwartz doesn't explicitly use the term *prototype* here, but as a price the term *core* is left unproblematized. And if he means *prototype*, then do we want to talk about the "core" being "underspecified," given that prototypes generally are *richly* structured? Schwartz's case study on the structure of *feminism* is ultimately about the details of *different people's* ideas of feminism: clearly extrapersonal. So it seems that ultimately the "underspecified core" is an extrapersonal structure. If so, again: do we want to use terms such as *prototype* and *radial category* when referring to it?

Pamela Morgan, in her studies on contested categories (Morgan 1992, 1998), seems to sense some problems with the Type-1/Type-2 distinction. In her 1992 study on *science* as a contested category, Morgan cites Lakoff's definition of radial category, which rules out categories where the central case is simply more general than its subcases. She then adds that "it may be argued that the noncentral cases of 'science' in the periods with only an underspecified, and not a prototype, core do not have any different properties from the general central case, and hence the type of categorization should be considered classical (based on necessary and sufficient conditions) and not radial" (Morgan 1992:5, fn. 11). As with Schwartz, however, the notion of core isn't problematized. Morgan hints at the need to problematize, though, by using two different modifiers with "core": "underspecified" and "prototype".

Remaining skeptical in her dissertation, Morgan (1998) notes that her case study, *propaganda*, could, when looked at two different ways, be seen as either Type 1 or Type 2. In her explanation, Morgan notes that the "underspecified core"—shared by all—"is then conventionally filled out with certain mainstream culturally acceptable external belief structures concerning politics, economics, and so on, so that in actual usage the prototypical example of 'propaganda' seems

rich, like ‘art’” (286-7). In other words, from the extrapersonal perspective, *propaganda* looks like Type 1; from the intrapersonal perspective, it looks like Type 2.

#### 4. Conclusion

By now, then, it should be apparent that we don’t need two types of contested category. If we look at any contested category from an extrapersonal perspective, it will look like Type 1, whereas from an intrapersonal perspective it will look like Type 2.

One question remains: what accounts for Lakoff’s original intuition that there are two different types of contested categories? Lakoff has explained (personal communication) that in early investigations, the cases of *democracy* and *art* seemed to behave quite differently, in that for *art* there was, after talking to enough people, nothing at all that was shared by everyone, whereas for *democracy* some schematic structure remained. From the perspective of this paper, the difference is again not a fundamental one of categories belonging to different “types,” but rather of categories that involve different degrees of sharedness being examined from two different angles: the intrapersonal and the extrapersonal.

#### References

- Connolly, William E. 1993/1974. *The Terms of Political Discourse*. 3<sup>rd</sup> edition. Princeton, NJ: Princeton University Press.
- Gallie, W. B. 1956. Essentially Contested Concepts. In *Proceedings of the Philosophical Society* 51:167-198. London: Harrison and Sons, Ltd.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Human Mind*. Chicago: University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, Volume I: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Morgan, Pamela S. 1992. ‘What is Science?’ and the Cognitive Theories of Contested Concepts and Prototypes. Ms., University of California, Berkeley.
- Morgan, Pamela S. 1998. *Beyond the Issues: A Linguistic and Conceptual Study of American Public Discourse*. Ph.D. diss., University of California, Berkeley.
- Patent, Jason D. 1997. Fairness, Discrimination, and Preferences: A Cognitive-Linguistic Investigation into the Conceptual Structures Underlying the Affirmative Action Debate. Ms., University of California, Berkeley.
- Rosch, Eleanor. 1975. Cognitive Reference Points. *Cognitive Psychology* 7:532-547.
- Schwartz, Alan. 1992. *Contested Concepts in Cognitive Social Science*. B.A. thesis, University of California, Berkeley.
- Shore, Bradd. 1996. *Culture in Mind: Cognition, Culture, and the Problem of Meaning*. New York: Oxford University Press.

*A Unified Account of Essentially Contested Concepts*

Strauss, Claudia, and Naomi Quinn. 1997. *A Cognitive Theory of Cultural Meaning*. Cambridge: Cambridge University Press.

Stanford Overseas Studies Program  
Shaoyuan Building 7, Room 309  
Peking University  
Beijing 100871 P.R. China

[jpatent@stanford.edu](mailto:jpatent@stanford.edu)



## Direct Quotation in Japanese: A Site for Stereotyping

KATSUE AKIBA REYNOLDS  
*University of Hawai'i at Manoa*

### 0. Introduction

Quotation of others' speech is one of the most common phenomena in the use of language. Discourses that we participate in as speakers/writers or hearers/readers, whether they are conversations, political speeches, conference papers, radio or TV news, novels, critiques, newspapers, magazines, or lectures, are full of quotations. Traditionally, quotations are identified either as "direct quotations" or as "indirect quotations", and it is generally assumed that a direct quotation is the exact reflection of what has actually been said. But how truthfully does a direct quotation reflect what has actually been said? Bakhtin (1981:340) observes that direct quotation is always subject to certain semantic changes no matter how accurately it is done. It is of great interest from a cognitive or sociolinguistic point of view to ask what the nature of the semantic changes is and how the changes occur. In this study I examine direct quotations in various Japanese texts, explore the mechanism of direct quotation, and demonstrate that direct quotation is the very site of stereotyping.

### 1. From Primary Utterance to Quotation

As preliminary observations, I first note the distinction between "primary speech genre" and "secondary speech genre" (Bakhtin 1986:62) and schematize the process of direct quotation as follows: Speaker A (primary speaker) says X (primary utterance), which expresses a set of meanings  $X_m$  (primary meanings); Speaker B (secondary speaker) interprets  $X_m$  as  $X_{m2}$  and attempts to encode  $X_{m2}$  into X2. In reality, there are cases in which the source of quotation is part of a non-conversational or written material, such as a public speech, a TV or newspaper news report, poetry, or a passage from a novel. In this exploratory study, however, I assume ordinary conversational interactions as the primary activity of language, and explore how semantic changes occur in the act of direct quotation. In other words, X is an utterance in "living conversation", while X2 may be part of a "living conversation" utterance or a written discourse.

Second, I wish to gain insight from the view that the signs of language, which are limited in number, must allow "built-in vagueness" in order to designate

reality, which is infinite (Bolinger and Sears 1981:110). Stereotyping occurs in the area of non-propositional meanings or pragma-sociolinguistic discourse meanings, which are in most cases expressed without the speaker's conscious intention. Such meanings tend to be communicated by paralinguistic means (or "audible gestures" and "visible gestures" in Bolinger and Sears's terms), and are more analogic, vague, and variable than propositional meanings. The secondary speaker fails to fully recover the primary meanings partly because there is some built-in vagueness in the primary utterance and because the intended meanings cannot be exactly encoded back into the primary utterance. It is more than likely that the primary meanings are to be changed to greater degrees than generally considered. Japanese provides interesting examples to show how great the changes can be. I examine four kinds of Japanese discourse involving direct quotation, and I demonstrate how stereotypes can occur.

## 2. Interactional Particles as "Audible Gestures"

Japanese is known as a language with an extremely elaborate honorific system and a number of "sentence-final particles", among other things. (Since the term "sentence-final particle" is misleading, as I pointed out elsewhere, I refer to these particles as "interactional particles", and when I mention sentence-final particles I mean "interactional particles occurring in the sentence-final position.") Although honorifics and interactional particles intimately and animately interact in a total communicative act, they are different in linguistic status; honorifics are basically part of the system of Japanese grammar, the honorific and the non-honorific being morphosyntactically distinguished, while the particles are on the boundary between grammar and paralinguistic. They convey more meanings by pronunciation than by the segmental form. They are "voice carriers." The intended meanings of particles cannot be clearly grasped without listening to the voice. Furthermore, particles carry different semantic values in different dialects (Reynolds 2000), which adds to the vagueness and complexity of the meanings to be conveyed by particles.

The verbal expressions *iki-masu* (go-POLITE) and *iku* (go-Ø) both convey the propositional information '(I) will go', but they are lexically differentiated; the former is accompanied by the honorific suffix *-masu* and the latter is not. The verbal forms are either "polite" or "non-polite." The meanings to be conveyed by particles, on the other hand, are not categorically specifiable, but they convey different kinds of experience with different degrees of emphasis depending on the way they are pronounced. Particles can be pronounced with different lengths and different intonations signaling complex messages as to the speaker's attitudes and feelings towards the hearer or the content of the propositional information. Utterances such as *ashita iku yo*, *ashita iku wa*, *ashita iku sa*, *ashita iku zo*, and *ashita iku ze*—in which the same proposition *ashita iku* ('tomorrow-go') is followed by a particle *yo*, *wa*, *sa*, *zo*, or *ze*—are all declarative statements meaning '(I) will go tomorrow'. The particles *yo*, *wa*, *sa*, *zo*, and *ze* added at the end do not affect the propositional meaning, but they are nevertheless an essential part of

### *Direct Quotation in Japanese*

the communicative activity of the Japanese. Particles play major roles in stereotyping. In order to make my point regarding stereotyping in the process of direct quotation, I need to discuss “potential meanings” (Bolinger and Sears 1981:109) without examining this heretofore unexplored territory in any great detail. What I give in the following are therefore schematic descriptions of the major interactional particles utilized for speech stereotypes. Interrogatives (e.g. *ka*, *kana*, and *kashira*) and particle combinations (e.g. *yo-na*, *yo-ne*, *wa-ne*, *wa-yo-ne*) are not included here to keep our discussion from becoming unnecessarily convoluted. (cf. Reynolds 1985 for more details.)

***Yo***: Expresses the speaker’s assertion with certainty and thus is used mainly by male speakers unless it combines with honorific verbs or other morphosyntactic adjustments as will be discussed below in connection with stylistic variations.

***Wa***: The weakest in the assertion. This particle pronounced with a rising intonation is a clear indication of the feminine identity of the speaker. It is confusing, however, because *wa* pronounced with a kind of falling intonation is used even more frequently by males than female speakers in western dialects such as the Osaka dialect.

***Sa***: Carries a tone of “strong insistence” and is said to be generally used by male speakers talking to their equals. However, it is rarely used as a sentence-final particle in contemporary conversations, as shown immediately below.

***Zo***: Conveys the speaker’s declarative and authoritative attitude and is used when the speaker has significantly greater power than the hearer does or when the speaker is doing a threatening act. It is therefore typically used by male speakers.

***Ze***: Similar to *zo*, but its meaning is more complicated than that of *zo*, perhaps because it is a fusion of *zo* and *i/e*, an older particle conveying the speaker’s intention to involve the hearer. This is also used exclusively by male speakers.

***Ne***: Expresses the speaker’s intention to involve the hearer in judging the validity of the propositional part of the statement. The statement followed by this particle pronounced with a rising intonation becomes close to an interrogative.

***Na***: Similar to *ne* except that this is more imposing than *ne*, and is said therefore to be used by male speakers. In my data, however, it is used by females as well.

It is clear that particles play crucial roles in stereotyping, especially in regard to men’s and women’s speech. There are strong indications, however, that particles are “voice carriers” and the meanings that their audible gestures convey are rich but vague. Some of the subtle meanings to be conveyed by particles are inevitably changed or lost in the process of transmission from the primary speech to the secondary speech, spoken or written.

### 3. Interactional Particles in Conversation

The first data concerning direct quotation is from the conversational data gathered by three Japanese sociologists for the purpose of their ethnomethodological study.<sup>1</sup> I chose 21 sets of their transcriptions, romanized them, and made a database for my own research of conversational analysis including interactional particles and pause fillers. (I did not use the sets from which some pages were missing.) The table in (1) below is a summary of the counting of interactional particles found in three different positions: sentence-final position, sentence-internal position, and sentence-external position (i.e. positions unrelated to the propositional structure of the sentence).

(1) Distribution of interactional particles

|          | <i>ne</i>   | <i>sa</i> | <i>yo</i>  | <i>na</i>  | <i>wa</i> | <i>zo</i> | <i>ze</i> | TOTAL       |
|----------|-------------|-----------|------------|------------|-----------|-----------|-----------|-------------|
| Final    | <b>1470</b> | <b>0</b>  | <b>445</b> | <b>116</b> | <b>9</b>  | <b>3</b>  | <b>2</b>  | <b>1817</b> |
| Internal | 1737        | 647       | 2          | 0          | 0         | 0         | 0         | 2384        |
| External | 74          | 0         | 0          | 0          | 0         | 0         | 0         | 74          |
| TOTAL    | 3281        | 647       | 447        | 116        | 9         | 3         | 2         | 4289        |

What is particularly relevant to this study of direct quotation is the fact that some particles occurred in great frequency, while others, especially those strongly gendered ones (*wa* for female speakers and *zo* and *ze* for males) were noticeably few. This limitation has recently been observed by Japanese researchers. Endo et al. (1989) found no instance of sentence-final *wa* in their natural conversation data (2056 utterances), and Ozaki (1994) found only one instance in the 734 samples gathered from natural conversations. Furthermore, most of *wa*, *zo*, and *ze* in my data occurred in direct quotations when speakers were mimicking others. A female student of a co-ed college, for example, says the following using *wa* in the final position of the quotation, imitating the feminine talk of women college students.

- (2) *watashi kekkon-suru wa.*<sup>2</sup>  
 I marry IP  
 ‘I will get married.’

<sup>1</sup> Three sociologists, Ehara Yumiko, Yoshii Hiroaki, and Yamazaki Keiichi, recorded and transcribed 32 sets of 30-minute conversations (10 male-male pairs, 10 female-female pairs, and 12 male-female pairs). I am very grateful to these researchers for their generous permission for the use of their transcripts for my analysis. Their data, methodically gathered without losing the naturalness of the setting, gave me much confidence in my findings.

<sup>2</sup> Abbreviations in the word-by-word glossing: COMP = complementizer (*to*), COP = copula (*de'aru/da*), COP=P = polite/formal copula (*desu*), DO = direct object (*o*), IP = interactional particle, NEG = negative, NOM = nominalizer (*no*), SUBJ = subject (*ga*), TOP = topic marker (*wa*).

Some comments on the result of the text count should be given at this point. First, the confirmative particle *ne* occurred more frequently in non-final positions than the final position. Second, *sa*, though it is described as a sentence-final particle in most studies of Japanese particles, did not occur at all as such in this data, while it occurred in internal positions with great frequency (647 instances). These are good reasons to consider it misleading to refer to all these particles as “sentence-final particles”. Another point to be brought up is the fact that *na*, which occurred quite frequently, needs a closer examination. In the majority of cases, *na* was pronounced with a prolonged vowel, namely as *naa* (although the tapes were not available to me, the prolongation was clearly indicated in the Japanese transcriptions), and 37 instances of the total 106 occurred in female speakers’ utterances. It is possible that particles that had come from different dialects are undergoing an overall change towards the formation of Standard Japanese, which began in the beginning of the modernization period of the history of Japan (i.e. sometime after the Meiji Restoration in 1868). I suspect that particles as used by dialect speakers or those which often appear as speech stereotypes in secondary genres, such as novels, popular songs, and TV dramas, may be mixed in the researchers’ intuitive data.

At any rate, the data suggest a curious relationship between the use of sentence-final particles and stereotyping in direct quotations. To see the relationship more clearly, I examined direct quotations in three types of written conversational discourse: “magazine interviews”, Japanese subtitles added to utterances of foreigners in TV news programs, and direct quotations in novels.

#### **4. Stereotyping in a Magazine Interview**

In Reynolds (1999), I compared the transcript of a TV interview of a Japanese woman (a primary genre text) and a “magazine interview” of the same woman (a secondary genre text). The woman, who was about 40 at the time, was confiding her sexual relationship with the then Prime Minister of Japan. First she talked with an editor of a popular weekly magazine and had him anonymously publish the story in the form of an interview, but she later appeared on TV to explain how it happened directly to the Japanese public. The difference between the TV interview and the magazine interview was striking.

The primary discourse is full of conversational features, such as incomplete sentences, pause fillers, afterthoughts, and instances of rephrasing and stammering. She talked decisively, assertively, hesitantly, or angrily depending upon the topic. She was very hesitant when she was asked to describe the scene in which the Prime Minister mentioned the price of her body, which she must have remembered as the most humiliating; she was determined and angry when she protested the comment that what had happened to her was just a question of the lower part of her body, not a political or social issue. It was evident on the TV screen that she was trying to appeal to the public with the utmost sincerity in spite of the difficulty that it was a question of human rights—a social issue. She was not performing the expected feminine role: there were no instances of the feminine

particle *wa* in her speech. In the magazine interview, on the other hand, an image of a woman with a different attitude was projected. Most of the 165 utterances in her responses were grammatically complete sentences ending with the polite finite form of the predicate, but with no hesitation. Of the 165 utterances, 80 were with a sentence-final particle. The table in (3) below is a summary of the 80 instances of particles.

(3) Particles used in the magazine interview

| <i>yo</i> | <i>ne</i> | <i>wa</i> | <i>mono</i> | other* | TOTAL |
|-----------|-----------|-----------|-------------|--------|-------|
| 43        | 22        | 5         | 1           | 9      | 80    |

\*“Other” includes *yone* (combination of *yo* and *ne*) and *ka* (interrogative).

Being dominated by the sense of finality of the complete sentences and assertion expressed by sentence-final particle *yo*, the magazine interview as a whole gave the impression that she was a woman who could talk about her own sexual matters without hesitation and in an imposing tone, in public. Many readers who were asked for comments after the publication of the magazine interview to be published in the subsequent issue of the magazine criticized her for being “nervy” and “shameless.”

### 5. Stereotyping Foreigners

The second example to show that meaning changes in direct quotation is a set of brief interviews of foreigners telecast on Japanese news programs. It is a recent trend to include primary utterances of foreigners in the TV report with Japanese subtitles. The audience viewing interviews with foreigners on TV news programs hear the primary utterance in a foreign language and/or read the secondary utterance on the screen.

It has been noticed that the subtitles are often grossly inconsonant with the primary utterances in terms of the impression of the speaker. The gap suggests a problem inherent to the translation of non-propositional meanings developed in the socio-historical reality of each culture. The producers (or translators) face the difficult task of overcoming the semantic gap between the two languages/cultures. They first interpret the primary utterance into what they believe is the Japanese equivalent of the primary utterance, applying the knowledge about the audible and visible signs conventionalized in the Japanese semiotic system; then, by applying the Japanese rules of meaning-form correspondence, they encode the interpreted meanings into readily transcribable forms to show on the screen. In an attempt to read non-propositional meanings which may not exist or are not important in the semiotic system of the culture of the primary speaker, they must make more recourse not only to “audible” messages, but also to “visible” messages and must come up with translations that they think would make sense within the Japanese culture. In other words, the semantic changes from the primary utterance to the

*Direct Quotation in Japanese*

secondary utterance in this case involve not only stereotyping but also “Japanization.”

I have analyzed 21 interviews telecast on FCI on January 1-2, 2001, in the two specials “Convergence: The Future TV News Programs Seen through the Eye of a TV Reporter” and “For Whom Are the Waves? Survival of Local Networks in the United States”. Each piece was analyzed in terms of the speaker’s gender, the professional status (affiliation and position), the attire (whether the clothes were formal, informal, or in between), and the language style as shown in the sentence-final mode in the subtitle. Even though the interviewees were all media-related professionals and their speech was not particularly different from each other, their utterances were translated in four different discourse styles. The table in (4) below schematically illustrates each style with a verbal predicate meaning ‘(I) will write’ consisting of a single verb in the middle column and a copular predicate meaning ‘(It) is Tokyo’ made up of *Tokyo* and a copula in the right column.<sup>3</sup>

(4) Four stylistic variations used in Japanese translations

| Style               | ‘(I) write/will write’         | ‘(It) is Tokyo’                   |
|---------------------|--------------------------------|-----------------------------------|
| I Neutral (written) | i. <i>kaku</i>                 | v. <i>Tokyo de’aru/ da</i>        |
| II Formal (spoken)  | ii. <i>kaki-masu (yo)</i>      | vi. <i>Tokyo desu (yo)</i>        |
| III Informal Male   | iii. <i>kaku (yo zo/ze...)</i> | vii. <i>Tokyo da (yo/zo/ze..)</i> |
| IV Informal Female  | iv. <i>kaku wa/wa yo</i>       | viii. <i>Tokyo yo/da wa yo</i>    |

The table in (5) below is a summary of the analysis of the 21 interviews in the order they appeared in the two programs. (Judgment of the attire is based on the observations of three people, two Americans and myself. When the turn of the same interviewee consists of more than one sentence, the earlier ones are sometimes translated with the neutral style and the last one or two in one of the other styles. In such cases, the style in sentences at the end characterizes the whole turn. The spellings of names have been inferred from the Japanese *katakana* writing and may not be accurate.)

---

<sup>3</sup> Styles are not of course determined solely by the sentence-final modes, but they are indicated by many other features throughout the discourse. The choice of first-person pronouns such as *watakushi*, *watashi*, *atashi*, *boku*, and *ore* is particularly relevant to the stylistic diversity.

Katsue Akiba Reynolds

(5) Interviewees on the FCI specials on January 1-2, 2000

| EX   | NAME              | M/F | POSITION             | ATTIRE  | STYLE |
|------|-------------------|-----|----------------------|---------|-------|
| 1-1  | Jason Austell     | M   | Reporter, KFMB       | -formal | III   |
| 1-2  | Bret Dumbron      | M   | Director, KFMB       | +formal | II    |
| 1-3  | ---               | M   | Technologist, WCYB   | -formal | II    |
| 1-4  | ---               | M   | Operator, WCYB       | -formal | III   |
| 1-5  | George Mazion     | M   | VP*, WCYB            | +formal | II    |
| 1-6  | Stacy Owen        | F   | Director, KRON       | -formal | IV    |
| 1-7  | Deborah McDermott | F   | VP, YB*              | +formal | II    |
| 1-8  | Bob Goldbenger    | M   | Director, KNTB       | ±formal | III   |
| 1-9  | Denis Walton      | M   | Representative, NAB* | +formal | II    |
| 2-1  | Meril Brown       | M   | VP, MSNBC.com        | -formal | II    |
| 2-2  | Michael Oreskes   | M   | Mod., ABC News*      | -formal | III   |
| 2-3  | Nicholas Barretta | M   | Pres., NEXTVENUE     | ±formal | III   |
| 2-4  | Jane Applegate    | F   | Pres., sbtv.com      | +formal | II    |
| 2-5  | Teresa Ferguson   | F   | VP, DirectTV         | +formal | I     |
| 2-6  | Tim Butcher       | M   | VP, WebTV.net        | -formal | II    |
| 2-7  | Michael Pope      | M   | Prof., MIT           | -formal | II    |
| 2-8  | Alex Bentland     | M   | Prof., MIT           | -formal | I     |
| 2-9  | Mitchel Lesnik    | M   | Prof., MIT           | -formal | I     |
| 2-10 | Richard Porpay    | M   | Prof., MIT           | -formal | II    |

(YB: Young Broadcasting, Mod: moderator, NAB: National Association of Broadcasting)

In the TV interviews, Style I, being unmarked for non-propositional meanings, is used to provide only the content of the propositional meaning. Style II conveys a sense of formality similar to the V-form in the theory of power semantics (Brown and Gilman 1960). It can be followed by interactional particles when used in face-to-face interactions. At the informal level the gender difference becomes apparent: in the male style, the non-polite verb can be followed by any declarative particle; in the female style, on the other hand, it can be followed by only *wa* or *wa-yo*, and various verbs need to be modified towards weaker assertion. The informal copula *da* of the copula nominal or adjectival predicate is typically suppressed in the female style as seen from a comparison between the male style *Tokyo da yo* and the female style *Tokyo yo* in the table in (4). Related to this non-use of the assertive copula *da* is the pattern [S *no*], which is commonly appropriated in the speech of female speakers. Although the *no* at the end of this construction is sometimes identified as a feminine sentence-final particle, from a formal point of view, it is a nominalizer. The [S *no*] is the zero-copula version of the [S *no desu/da*] construction and is roughly translated into English with ‘it is that S’ (Kuno 1973:223-233). Compare (6) and (7).

*Direct Quotation in Japanese*

- (6) *unzari shite-i-ta no. . . . . wakuwaku suru wa.*  
disgusted be-PAST NOM excited be IP  
'(It is that) I have been disgusted. . . . (I'm) excited.'
- (7) *isogashiku naru yo. san-ji kara bangumi ga hajimaru n da.*  
busy become IP 3-o'clock from program SUBJ begin NOM COP  
'(We will) become busy. It is that the program will begin at three o'clock.'

Example (6) consisting of [S *no*] and [S *wa*] is from the speech of Stacy Owen, a female news director of KRON, and (7) with [S *yo*] and [S *n da*] (*no* being contracted) is from the speech of a male staff member of WCYB. Since Japanese counterparts of these professionals would never talk in these informal styles in similar contexts, it is curious why and how the Japanese translators make such stylistic distinctions. Audible and visible gestures as well as social variables, such as gender, age, and social status, seem to be important clues.

If the speaker is a senior male with a distinguished title, wearing a formal coat and tie, the subtitle would be in Style II (e.g. 1-2, 1-5). If the speaker is a male speaker in casual work clothes, it will be in Style III (e.g. 1-1, 2-2). A female media person is interpreted as a user of the formal spoken style if she is wearing, for example, a jacket with a necklace on (e.g. 1-7, 2-4), but if she is relatively young and wearing casual clothes, she is likely to be represented by the informal women's style as in (6), in which all the assertive force of the utterance is reduced or eliminated. There are of course cases in which perceived semiotic indicators are contradictory or confusing to the interpreter. For example, sample 2-3 is a male company president wearing a coat and tie, but his utterances were translated by a mix of Style I and Style III, suggesting that the translator somehow sensed a relatively strong maleness in him although he was by other standards a person who would talk in Style II. It may be his fast speech with three instances of quite noticeable stammering (*they will | they | they are...*, *in | in terms of...*, *it | it's...*), three quick pauses and four instances of the pause filler *ah* within the 46-second turn, or sharp penetrating eyes and boldness in the front—namely, “visible gestures”—that made the translator perceive him as an aggressive masculine man. Sample 2-9 was an MIT professor wearing a sweatshirt instead of a formal coat and tie, which is expected of a prestigious university professor in Japan. His utterances were translated in Style I, the unmarked style.

The examples in these programs are not particularly special. I have observed these news programs for years, transcribing many instances showing significant semantic changes from the primary to the secondary. A farmer in South America, men on the street in Israel after the presidential election, and male earthquake victims in El Salvador were all represented in the informal male speech style. The distressed and angry shouts of a Russian woman and mother of a crew member who died in a submarine accident, protesting to a Russian official, were translated with informal women's style. The anger expressed not only in her voice but also in her whole body and the Japanese female language style was an extremely

awkward mismatch. The female language style is for Japanese women, who are not supposed to ever show anger.

The stereotyping of foreigners' discourse according to Japanese ideology is a serious problem emerging as Japanese TV programs become increasingly internationalized. During the Atlanta Olympics, a university teacher wrote a letter to the editor of the magazine *Asahi*, pointing out that Canadian gold medalist Donovan Bailey, who "spoke in a voice with confidence and with humility" in what the college teacher perceived as proper English, "even sounding like a thoughtful philosopher," was stereotyped as an "unsophisticated, coarse male" when his interview was translated with the informal male Japanese style. The author of the letter wrote, "The translation is awfully unpleasant to me: the translator is too insensitive...Isn't this a form of racism?" Yabe (1996) examined all the direct quotations (marked by quotation marks) that appeared in the three major dailies during the Atlanta Olympics: all the Japanese athletes, men and women, responded with the formal polite style (Style II), while foreigners' utterances were translated with different styles, female athletes' comments with the informal female speech style and the comments by male athletes, particularly non-whites, were translated with the informal masculine style.

## 6. Creating Novel Characters

The third place where one can see the gap between primary and secondary genre speech is the novel. Bakhtin defines the novel as an artistically organized diversity of social speech types and individual voices, and states:

The internal stratification of any single national language into social dialects, characteristic group behavior, professional jargons, genetic languages, languages of generations and age groups, tendentious languages, languages of authorities, of various circles and of passing fashions, languages that serve the specific sociopolitical purposes of the day, even of the hour...this internal stratification present in every language at any given moment of its historical existence is the indispensable prerequisite for the novel as a genre. (Bakhtin 1981:262-3)

The diversity is most concretely observed in the characters' dialogues, which are direct quotations in the sense that they represent the voices of others in a fictitious world. Characters' dialogues can be far more extreme in stereotyping because of the inevitable loss of voice in the process of going from spoken to written discourse and because of the stylization chosen by the author, abstracting some relevant characteristics of social and individual speech types, to be maximally effective for the artistic purposes. To support this point, I have analyzed four major characters of *Hard-Boiled Wonderland and The End of the World* (*Wonderland*, hereafter), a novel of nightmare fantasy by Murakami Haruki (1949-).<sup>4</sup>

---

<sup>4</sup> Murakami is known in the West through English translations, such as *The Wind-Up Bird Chronicle* and *Norwegian Wood*.

### Direct Quotation in Japanese

The characters of this twilight zone are deprived of their shadows and do not have any memory of their past. None of the characters are given personal names. They are referred to by second-person pronouns (e.g. *anata*, *anta*, *omae*, and *kimi*) in the dialogues and by their categories (e.g. “Old Man,” “Gatekeeper,” “Fatty Woman in Pink Dress,” and “Librarian Girl”) or by the gendered third-person pronouns *kare* (male) and *kanojo* (female) in the narratives. More interesting is the final mode where detailed interactional meanings are created through interactions of various linguistic dimensions, such as formal vs. non-formal verbals, classical vs. contemporary lexical items, and the presence vs. non-presence of interactional particles. I have examined the first 100 utterances of each of the four characters talking to the author-narrator and have conducted a text count analysis of their final modes.

The “old man” character, who hired the narrator for his research purpose, is a biologist with some connection with “the central system” of the Wonderland. His utterances are consistently in Style II, suggesting his seniority and status. The method of achieving this stylistic level is not by the standard formal verb suffix, but by the [S *desu*] construction, a somewhat deviant variety of the [S *no desu*] construction. He says *kaku desu* (‘I will write’) instead of *kaki-masu* or *kaku no desu*. 39 of the 100 samples are in this peculiar construction, 7 in the standard [S *no desu*] construction, and 3 in the [S *wake desu*] construction (*wake* being a variant of the nominalizer). There are only 3 instances of verbals with *-masu*. What is more peculiar in the speech of the biologist is the particle *na* following the *desu* of the [S *desu*] construction and the *desu* of the regular nominal or adjectival predicate, as seen in (8) below.

- (8) *watashi wa zukotsu o atsumete-oru desu na.*  
I TOP skull DO collect-BE COP IP  
‘I have been collecting skulls.’

All 14 instances of *na* follow *desu*, thus making the utterance-final mode *desu na* a unique feature of this character’s speech. The interactional particle *zo*, the one with the greatest degree of assertion, combined with *desu* is also found in the speech of the same character, though not as frequently as the mode *desu na*.

Final modes, such as *desu na* and *desu zo*, may sound unusual and contradictory since *desu* conveys formality, politeness, distance, etc., and the particles *na* and *zo* carry the opposite semantic values, such as masculine intimacy or intention to involve the hearer, but they actually occur in the speech of some male speakers. The appropriation of this mode as a feature of the biologist’s speech is not totally unrealistic. It is the frequency of this mode in his speech that is unrealistic and responsible for stereotyping.

The speech of the gatekeeper to the Wonderland is almost the opposite of the biologist’s. He is depicted as a virile man of lower status. He talks exclusively in Style III, constantly using the informal copula *da* or masculine particles *sa* and *na* as shown in (9)-(11) below.

- (9) *anta mo kawatta hito da na.*  
 you also different person COP IP  
 ‘You are also a strange person.’
- (10) *furui yume o yomu n da.*  
 old dreams DO read NOM COP  
 ‘(You) will read old dreams.’
- (11) *wakaru sa.*  
 understand IP  
 ‘(I) understand (you).’

What [*S desu*] is to the biologist is what [*S n da*] is to the gatekeeper. 20 of the 100 samples of the gatekeeper’s speech are in this form. Counting the utterances ending with the *da* of the regular copular predicate, 55 end with *da*, with the particle of strong assertion *yo* (3 cases), the particle of masculine intimacy *na* (12) or *sa* (9), or with no particle. The speech of the gatekeeper, totally lacking the formality required in talk with a stranger, gives the impression that he is “uncivilized.”

The speech of the female characters is not very different from person to person. Both the librarian and the biologist’s granddaughter talk in the informal female style (Style IV). 21 of the first 100 utterances of the librarian’s speech end with *wa*. Also salient is the frequent use of [*S no*], the zero-copula version of the [*S no da*] construction. 42 of the 100 samples are in this form. In addition, the copula of the regular copula sentence is always left out as in (12) below.

- (12) *koko wa totemo shizukana machi yo.*  
 this.place TOP very quiet town IP  
 ‘This is a very quiet town.’

The biologist’s granddaughter’s speech is the same as the female librarian’s although she is much younger.

There are some other linguistic devices with which the author attempts to achieve his intended image of *Wonderland*. What I have shown, however, must be sufficient to see how only certain meanings are disproportionately expanded and how grossly characters are stereotyped.

## 7. Conclusions

In this study, primary genre data and several sets of secondary genre data were comparatively analyzed in terms of the semantic changes in the process of direct quotation. The data amply demonstrated that semantic changes in the transmission of another’s speech are inevitable, and it is indeed true that “we stereotype” (Lakoff 1990:181). We are constantly exposed to speech stereotypes when we converse with others, read magazines, watch TV, and read novels, with little

awareness of what is happening, resulting in unconscious collaboration with the stereotyping. Japanese linguists, including myself, have attempted to describe the pragma-sociolinguistic meanings of interactional particles and other ingredients of stereotyping, taking primary conversational data and secondary data indiscriminately all together. Consequently we have not only been prevented from an adequate understanding of those areas of the language, but we have also helped further enforce existing stereotypes. I wish to emphasize here that as the college teacher wrote to *Asahi*, stereotyping can be harmful and it functions as a conservative force preserving old ideology. The media needs to be more sensitive to the socio-psychological ramifications of their language use instead of continuing to indulge in existing stereotypes. Even novels should not be completely exempt from the accusation of stereotyping merely on the grounds that it is necessary for artistic creation. Several novels by Murakami Haruki including *Wonderland* have been translated into English such that most of the meanings conveyed by speech stereotypes in the Japanese original have been lost, but the value of the novels is nevertheless appreciated. The book reviewer of *The Atlantic*, for example, has commented that *Wonderland* is a provocative work “rich in action, suspense, odd characters and unexpected trifles.” Also, there are younger Japanese writers who are creative in reflecting in characters’ speech the emerging linguistic diversity in contemporary Japanese. Language abounds in resources to create styles better suited to equality and globalization. Linguistics can suggest ways that stereotypes are produced and can show the harmful consequences of categorization.

## References

- Bakhtin, M. M. 1981. *The Dialogic Imagination*. Austin: University of Texas Press.
- Bakhtin, M. M. 1986. The Problems of Speech Genres. In Caryl Emerson and Michael Holquist (ed.), *Speech Genres and Other Late Essays* (trans. Vern W. McGee). Austin: University of Texas Press, 60-102.
- Bolinger, D. L., and D. A. Sears. 1981. *Aspects of Language*. New York: Harcourt Brace Jovanovich.
- Brown, R., and A. Gilman. 1960. The Pronouns of Power and Solidarity. In T. A. Sebeok (ed.), *Style in Language*. New York: John Wiley, 253-276.
- Chafe, Wallace L. 1982. Integration and Involvement in Speaking, Writing and Oral Literature. In Deborah Tannen (ed.), *Spoken and Written Language: Exploring Orality and Literacy*. Norwood, NJ: ALEX, 35-53.
- Ehara, Yumiko, Yoshii Hiro’aki, and Yamazaki Keiichi. 1993. Seisabetsu no Esunomesodorogii [An Ethnomethodological Study of Sex Discrimination]. In Reynolds (ed.), *Onna to Kotoba [Women and Language]*. Tokyo: Yushindo, 189-228.
- Endo, Orié, et al. 1989. Josei no Hanashi Kotoba—Terebi no Intavyuu Bangumi kara [Women’s Spoken Japanese—From Interviews on TV]. *Kotoba* 17.

Katsue Akiba Reynolds

- Kuno, Susumu. 1973. *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.
- Lakoff, Robin. 1990. *Talking Power*. New York: Harper Collins.
- Ozaki Yoshimitsu. 1994. Shuu Joshi ni-tsuite: Datsu-Josei-Sen'yoo-Go no Keikoo [On Sentence-Final Particles: Tendency towards De-genderization]. A Report of the Tokyo Josei Zaidan Grant Project.
- Reynolds, Katsue Akiba. 1985. Female Speakers of Japanese. *Feminist Issues* 5(2):13-46.
- Reynolds, Katsue Akiba. 1999. Discourse and Social Change. *Kotoba* 20:18-34.
- Reynolds, Katsue Akiba. 2000. A Sociohistorical Perspective: Implications of a Quantitative Analysis of Japanese Interactional Particles. *Kotoba* 21:88-103.
- Yabe, Hiroko. 1996. Shinbun Hodo no Gaikokujin Danwa ni Miru Dan-jo-sa [The Gender Difference in Foreigners' Speech Observed in Newspaper Reports]. *Kotoba* 17:58-72.

University of Hawai'i at Manoa  
Department of East Asian Languages and Literatures  
1890 East-West Road, Moore 361  
Honolulu, HI 96822

katsue@hawaii.edu

## Valence, Voice, and Noun Incorporation in Karajá

EDUARDO RIVAIL RIBEIRO

*University of Chicago*

### 0. Introduction

The purpose of this paper is twofold.<sup>1</sup> First, it provides a brief description of the morphosyntax of Karajá, a Macro-Jê language spoken in Central Brazil, focusing especially on the mechanisms of valence marking and grammatical relation changing. In addition, this paper discusses the implications of the Karajá data for Baker's (1988) incorporating account of antipassive. Karajá is traditionally described as having a very irregular fused set of prefixes indicating person, aspect, object, and direction (Fortune and Fortune 1964, Wiesemann 1986, Maia 1998). However, a more careful analysis reveals a rather regular, mostly agglutinating morphology, with separate prefixes indicating *person* (and, cumulatively, *mood*),

---

<sup>1</sup> I would like to thank the Wenner-Gren Foundation for Anthropological Research, the University of Chicago, and the Brazilian National Scientific Development Council, CNPq (Grant 200018/98-1), for their financial support. I also owe special thanks to the Karajá speakers for kindly teaching their language to me and to Neha Dave and Victoria Marty for their encouragement. Amy Dahlstrom provided valuable suggestions on an earlier version of this paper, for which I am very much indebted. Any remaining mistakes are, of course, my sole responsibility.

Karajá is spoken along the Araguaia River, in the states of Goiás, Mato Grosso, Tocantins, and Pará. It has four mutually intelligible dialects: Southern Karajá, Northern Karajá, Javaé, and Xambioá. The language presents systematic differences between male and female speech. Female speech can be postulated as the basic one. Male speech is characterized by the deletion of a velar stop occurring in the corresponding female speech form (e.g. ♀ *ɗkarã* 'I' > ♂ *ɗarã*). Unless otherwise noted, the data in this paper are from the female speech of the Southern and Northern Karajá dialects, but the grammatical features presented here are common to all four dialects. The data, obtained from native speakers in several fieldtrips, are presented both in phonetic (first line, in italics) and phonological transcription (second line). For details on the phonology, see Ribeiro (2000a).

Abbreviations and symbols: ♀ = female speech, ♂ = male speech, ADM = admonitory, AL = allative, ANTI = antipassive, CAUS = causative, CTFG = centrifugal direction, CTPT = centripetal direction, FUT = future, IMPERF = imperfective, INSTR = instrumental, INTR = intransitive, LOC = locative, NARR = narrative particle, PASS = passive, PERF = perfective, PROGR = progressive, REFL = reflexive, REL = relational prefix, TRANS = transitive, VERB = verbalizer. Roman numerals indicate the formal class to which the verb or noun stem belongs (see §2.1).

*direction*, and *valence*, e.g. (1). In addition, pronominal direct objects are obligatorily incorporated into the verb, e.g. (2).

- (1) *dɪkarẽ waha karɪrakokre*  
 dɪkarẽ waha ka-r-ɪ-rakɔ=kəre  
 I my.father 1-CTFG-TRANS-wait=FUT  
 ‘I will wait for my father.’

- (2) *waha rɪwarakokre*  
 waha Ø-r-ɪ-wa-rakɔ=kəre  
 my.father 3-CTFG-TRANS-1-wait=FUT  
 ‘My father will wait for me.’

These prefixes include a well-developed system of voice and valence markers. There is a clear-cut morphological differentiation between transitive and intransitive verbs, and valence changes are always morphologically indicated. Furthermore, the language presents a complex set of mechanisms to indicate changes in the grammatical relations among the arguments of a verb. For example, any transitive verb, such as *kəɔ* ‘to cut’, e.g. (3a), may be inflected to indicate the suppression of the agent, in a *passive* construction, e.g. (3b), or the suppression of the patient, in an *antipassive* construction, e.g. (3c).

- (3) a. *nadɪ waritʃɔre rade rɪkrɔrɛɪ*  
 d-adɪ wa-rikɔre rade Ø-r-ɪ-kəɔ=r-ɛɪɪ  
 REL-mother 1-offspring hair 3-CTFG-TRANS-cut=CTFG-PROGR  
 ‘My mother is cutting my child’s hair.’
- b. *waritʃɔre rade rakɔrɛɪ*  
 wa-ritʃɔre rade Ø-r-a-kəɔ=r-ɛɪɪ  
 1-offspring hair 3-CTFG-PASS-cut=CTFG-PROGR  
 ‘My child’s hair is being cut.’
- c. *nadɪ rɔkrɔrɛɪ*  
 d-adɪ Ø-r-ɔ-kəɔ=r-ɛɪɪ  
 REL-mother 3-CTFG-ANTI-cut=CTFG-PROGR  
 ‘My mother is cutting (something).’

The language also presents *noun incorporation*, which is generally a mechanism of possessor raising. In general, only inherently possessed nouns, such as body part terms, may be incorporated. Since only the possessed noun is incorporated, the valence of the incorporating verb remains unchanged, as the possessor is promoted to subject with intransitive verbs, or direct object with transitive

verbs, e.g. (4a). And since a transitive verb remains transitive after having incorporated a noun, it can still be made passive, e.g. (4b), or antipassive, e.g. (4c):

- (4) a. *nadı*      *waritʃɔɛ*      *rɪradɛkrɔɛɪ*  
d-adɪ      wa-ritʃɔɛ      Ø-r-ɪ-radɛ-kəɔɔ=r-ɛɪɪ  
REL-mother   1-offspring   3-CTFG-TRANS-hair-cut=CTFG-PROGR  
‘My mother is cutting my child’s hair.’  
[lit., ‘My mother is hair-cutting my child.’]
- b. *waritʃɔɛ*      *raradɛkrɔɛɪ*  
wa-ritʃɔɛ      Ø-r-a-radɛ-kəɔɔ=r-ɛɪɪ  
1-offspring   3-CTFG-PASS-hair-cut=CTFG-PROGR  
‘My child’s hair is being cut.’  
[lit., ‘My child is being hair-cut.’]
- c. *dadı*      *rɔradɛkrɔɛɪ*  
d-adɪ      Ø-r-ɔ-radɛ-kəɔɔ=r-ɛɪɪ  
REL-mother   3-CTFG-ANTI-hair-cut=CTFG-PROGR  
‘My mother is cutting hair.’  
[lit., ‘My mother is hair-cutting (someone).’]

Examples such as (4c), in which antipassive markers can co-occur with an incorporated noun, may have interesting implications for theories that treat antipassive as a special kind of noun incorporation, such as seen in Baker’s (1988) approach. If antipassive is merely a case of noun incorporation, as Baker claims, how would it interact with noun incorporation proper? The following section describes the morphosyntax of Karajá, in order to familiarize the reader with the mechanisms of grammatical relation changing in this language. Section 3 discusses the implications of the Karajá data for Baker’s account of the antipassive construction.

### 1. Morphosyntactic Overview

Karajá is an SOV head-marking language. Core NP arguments—that is, subject and direct object—are not morphologically marked. Pronominal subjects are expressed by a series of free pronouns, such as *dɪkarɔ̃* ‘I’ in (1) above, while pronominal objects are expressed by a series of bound morphemes, such as *wa-* ‘1<sup>st</sup> person’ in (2). In contrast to a fairly simple nominal morphology, Karajá presents a complex verb morphology. Stem formation processes such as compounding, noun incorporation, and reduplication are commonly used. In addition, as mentioned before, the verb is marked for *person* (and, cumulatively, *mood*), *direction*, and *voice/valence*.

**1.1. Lexical Classes**

Most noun and verb stems can be divided into two lexical classes, arbitrarily labeled *Class I* and *Class II* (Ribeiro 1996). The main difference between Class I and Class II noun stems is in the series of personal prefixes they take, as illustrated below by the paradigms for *kɔɾʊ* I ‘forehead’ in (5) and *ɛbɔ* II ‘hand’ in (6). The series of personal prefixes occurring with Class I and Class II stems are summarized in Table 1 below.

- |     |                                                                                                                                                                                                                                      |     |                                                                                                                                                                                                     |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (5) | <p><u>Class I</u></p> <p><i>habu kɔɾʊ</i> ‘man’s forehead’</p> <p><i>wa-kɔɾʊ</i> ‘my forehead’</p> <p><i>a-kɔɾʊ</i> ‘your forehead’</p> <p><i>ɪ-kɔɾʊ</i> ‘his/her/its forehead’</p> <p><i>da-kɔɾʊ</i> ‘his/her/its own forehead’</p> | (6) | <p><u>Class II</u></p> <p><i>habu d-ɛbɔ</i> ‘man’s hand’</p> <p><i>wa-d-ɛbɔ</i> ‘my hand’</p> <p><i>∅-ɛbɔ</i> ‘your hand’</p> <p><i>d-ɛbɔ</i> ‘his/her/its hand’, or<br/>‘his/her/its own hand’</p> |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

**Table 1.** Possessive prefixes in Karajá (Ribeiro 1996)<sup>2</sup>

| Person               | Class I    | Class II   |
|----------------------|------------|------------|
| 1 <sup>st</sup>      | <i>wa-</i> | <i>wa-</i> |
| 2 <sup>nd</sup>      | <i>a-</i>  | <i>∅</i>   |
| 3 <sup>rd</sup>      | <i>ɪ-</i>  | <i>d-</i>  |
| 3 <sup>rd</sup> REFL | <i>da-</i> |            |

Whereas the Class I prefix series distinguishes a reflexive third person (*da-*) from a non-reflexive one (*ɪ-*), the Class II series has only one third-person prefix (*d-*), which covers the range of meanings of both reflexive and non-reflexive third persons. Furthermore, the Class II stem *ɛbɔ* ‘hand’ presents a prefix *d-* in the first person and when preceded by a nominal possessor.<sup>3</sup> The function of this prefix is synchronically fairly opaque, but its distribution resembles that of *relational prefixes*, which mark the contiguity or non-contiguity of a stem to its determiner. Relational prefixes were first described as a grammatical peculiarity of Tupí-Guaraní languages, and their occurrence in languages of Karíb and Macro-Jê stocks, as well as in languages of other branches within the Tupí stock, has been pointed out as evidence for a genetic relationship among these three groups (Rodrigues 1994).

The main difference between Class I and Class II verb stems is in the fact that Class II intransitive verbs, such as *udədə* II ‘to become cold’, are marked by a

<sup>2</sup> Although all Class II stems are vowel-initial and most Class I stems are consonant-initial, the distinction cannot be reduced to phonological terms, since Class I also includes some vowel-initial stems, such as *aθikɔ* ‘arm’ (*∅ aθiɔ*), *εθɔ* ‘cotton’, *ari* ‘to gather’, *eθi* ‘younger brother’, etc.

<sup>3</sup> The relational prefix and the third-person prefix are palatalized before [high, +ATR] vowels, being pronounced as [dʒ] and [tʃ], respectively (cf. Ribeiro 2000a). In addition, the relational prefix has two non-phonologically conditioned allomorphs, *d-* (e.g. *d-ɛbɔ* ‘hand’) and *l-* (e.g. *l-ɔahr* ‘medicine’), whose choice is subject to dialectal variations.

zero prefix, whereas Class I intransitive verbs, such as *dɔdɔkɛ* I ‘to become hot’, are marked by the prefix *a-*.<sup>4</sup>

- (7) a. *bədə rudɔdɔra*  
*bədə Ø-r-Ø-udɔdɔ=r-a*  
 weather 3-CTFG-INTR-become.cold=CTFG-PERF  
 ‘The weather got cold.’
- b. *bədə radɔdɔkɛra*  
*bədə Ø-r-a-dɔdɔkɛ=r-a*  
 weather 3-CTFG-INTR-become.hot=CTFG-PERF  
 ‘The weather got hot.’

Class II transitive stems, such as *ʋahɪdɔ̃* II ‘to treat’,<sup>5</sup> take the prefix *d-* when not immediately preceded by a pronominal direct object or by an incorporated noun, e.g. (8a). When immediately preceded by a pronominal direct object or an incorporated noun, the relational prefix *l-* is used, e.g. (8b):

- (8) a. *hʌri waritʃɔɾɛ rɪdʋahɪnɔ̃rɛ*  
*hʌri wa-ritʃɔɾɛ Ø-r-I-d-ʋahɪ-dɔ̃=r-e*  
 shaman 1-child 3-CTFG-TRANS-3/REL-medicine-VERB=CTFG-IMPERF  
 ‘The shaman treated my child.’
- b. *hʌri rɪwalʋahɪnɔ̃rɛ*  
*hʌri Ø-r-I-wa-l-ʋahɪ-dɔ̃=r-e*  
 shaman 3-CTFG-TRANS-1-REL-medicine-VERB=CTFG-IMPERF  
 ‘The shaman treated me.’

## 1.2. Subject Agreement

Person agreement displays a strictly nominative pattern, with the verb always agreeing with the subject, be it intransitive (9a) or transitive (9b). Person agreement markers are distributed into two different sets, one occurring in the realis

<sup>4</sup> Maia (1998:28) terms the vowel that follows the directional prefix a *thematic vowel*, following an infelicitous tradition that traces back to Fortune and Fortune (1964). However, as we have seen, these vowels may be inflectional prefixes, marking voice and valence, e.g. (1)-(4), as well as person, e.g. (11). They can also be the result of the fusion between a person prefix and a voice prefix, e.g. (15a). Finally, they can be simply part of the verb stem, such as in the third person of Class II intransitive stems, e.g. (7a), (12a), (14a).

<sup>5</sup> This is a denominal verb derived from *ʋahɪ* II ‘medicine’ (cf. §1.4.1.1). It is tempting to consider the prefix *d-* occurring with transitive verbs as simply a marker of third-person object. However, this prefix also occurs with *antipassive* constructions, e.g. (20), which do not allow explicit direct objects (see §1.4.2.3).

(present and past tenses) and the other in the irrealis (future, potential, and admonitory). These prefixes are listed in Table 2 below.<sup>6</sup>

**Table 2.** Subject agreement markers in Karajá (Ribeiro 1996)

| Person          | Realis     | Irrealis                     |
|-----------------|------------|------------------------------|
| 1 <sup>st</sup> | <i>a-</i>  | ♀ <i>ka-</i> / ♂ <i>a-</i>   |
| 2 <sup>nd</sup> | <i>dá-</i> | <i>bə-/b-</i>                |
| 3 <sup>rd</sup> | ∅          | ∅; ♀ <i>k-</i> / ♂ <i>-ə</i> |

- (9) a. *kraritifakre*  
*ka-r-a-rika=kəre*  
 1-CTFG-INTR-walk=FUT  
 ‘I will walk.’
- b. *kariθuhokre*  
*ka-r-ɪ-∅-θuhə=kəre*  
 1-CTFG-TRANS-3-wash=FUT  
 ‘I will wash it.’

### 1.3. Direction

The verb also inflects for *direction*, according to the speaker’s point of view. *Centrifugal* direction (‘thither’), marked by the prefix *r-* or by its zero-allomorph, indicates that the process is seen as occurring away from the current location of the speaker, e.g. (10a). *Centripetal* direction (‘hither’), marked by the prefix *d-*, indicates that the process is seen as occurring towards the current location of the speaker, e.g. (10b). Centrifugal direction is the unmarked member of the opposition. All verbs are marked for direction, including those that apparently do not indicate a motion whatsoever, such as *uvu* ‘to die’ (cf. Ribeiro 2000b). Notice that the clitic aspectual auxiliaries also inflect for direction (cf. (3), (4), (7), (8) above), and, in the second person, also for person (cf. (15)).

- (10) a. *kariwikre*  
*ka-r-ɪ-∅-wi=kəre*  
 1-CTFG-TRANS-3-carry=FUT  
 ‘I will take it.’
- b. *kadiwikre*  
*ka-d-ɪ-∅-wi=kəre*  
 1-CTPT-TRANS-3-carry=FUT  
 ‘I will bring it.’

### 1.4. Valence and Voice

Karajá verbs are lexically either transitive or intransitive.<sup>7</sup> Intransitive verbs may have their valence increased through causativization (cf. §1.4.1.1) or through

<sup>6</sup> The same set of prefixes is used for singular and plural. There is also a distinction between a first person plural exclusive (marked by the same set of prefixes used for first person singular) and a first person plural inclusive (inflected for third person). The prefix *kə-* ‘3<sup>rd</sup> person’ is restricted to the centripetal direction of the irrealis mood.

<sup>7</sup> Maia (1998:79) mentions the existence of ‘diffuse verbs’—that is, verbs that can be used either transitively or intransitively without any morphological alteration. In our data, however, the only verb he mentions as being diffuse, *ukə* II ‘to dry’, has exactly the same behavior of other intransitive verbs, such as *uka* II ‘to be cooked’, e.g. (12). As shown in the example (b) below, this verb presents transitive morphology when used transitively. Notice that the transitive stem is a denomi-

oblique promotion (cf. §1.4.1.2). Transitive verbs, on the other hand, may have their valence decreased through reflexivization (cf. §1.4.2.1), passivization (cf. §1.4.2.2), and antipassivization (cf. §1.4.2.3).

### 1.4.1. Intransitive Verbs

Intransitive verbs are those that do not take a direct object as one of their arguments, such as *udədə* II ‘to become cold’, cf. (7a); *dədəke* I ‘to become hot’, cf. (7b); and *rika* I ‘to walk’, cf. (9a). As we have seen above, Class I intransitive verbs are generally marked by the prefix *a-*, while Class II intransitive verbs are marked by a zero allomorph. In addition, a few intransitive verbs, such as *θe* ‘to fall’, are marked by the prefix *ε-*. The class of intransitive verbs includes not only one-place verbs such as *rika* I ‘to walk’ and *udədə* II ‘to become cold’, but also *pseudo-transitive* verbs such as *obi* II ‘to see’, whose arguments are oblique NPs—in this case, a locative, marked by the postposition *bã* ‘diffuse locative’, e.g. (11). Although notionally transitive, such verbs behave as intransitive for all purposes. For example, they cannot be made passive or antipassive, and their arguments cannot be incorporated (16b).

- (11) *dɪkarẽ haləkɔɛmẽ rabire*  
 dɪkarẽ haləkɔɛ=bã r-a-Ø-obi=r-e  
 I jaguar=LOC CTFG-I-INTR-see=CTFG-IMPERF  
 ‘I saw the jaguar.’

Most intransitive verbs can be transitivized, either through causativization or through the promotion of an oblique to direct object. The transitivized stem is formed by the nominal form of the verb plus the verbalizer suffix *-dã*. This is illustrated in the example (12b) below, where the intransitive verb *ɔka* II ‘to be cooked’ is transitivized:<sup>8</sup>

---

nal verb formed by the deverbal noun *urã* ‘the action of drying’ followed by the verbalizer suffix *-dã* (cf. §1.4.1):

- |                                                                                                                                            |                                                                                                                                                                                |
|--------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(a) <i>bεε rukãre</i><br/>         bεε Ø-r-Ø-ukã=r-e<br/>         water 3-CTFG-INTR-dry=CTFG-IMPERF<br/>         ‘The water dried.’</p> | <p>(b) <i>tʃuu bεε rɪdʊrãnãkre</i><br/>         dʊ bεε Ø-r-I-d-urã-dã=kɔre<br/>         sun water 3-CTFG-TRANS-3/REL-dry-VB=FUT<br/>         ‘The sun will dry the water.’</p> |
|--------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

<sup>8</sup> This example illustrates a very common process for deriving nouns from verb roots, namely *consonantal replacement*, which consists in replacing a velar stop or a glottal fricative occurring in the last syllable of the verb root with an alveolar flap: *rika* I ‘to walk’ > *rira* ‘the action of walking’, *rira-da* ‘walking place’, *rira-du* ‘the one who walks’; *θuhɔ* I ‘to wash’ > *θurɔ* ‘the action of washing’, *θurɔ-da* ‘washing material’, *θurɔ-du* ‘the one who washes’ (Ribeiro 1996). Thus, the transitive stem in (12b) above is constructed with the nominal form of the verb *ɔka* ‘to be cooked’, *ura* ‘the action of cooking’, followed by the verbalizer suffix *-dã*.

- (12) a. *iweru rukareri*  
*iweru* Ø-r-Ø-*uka*=r-eri  
 calugi 3-CTFG-INTR-be.cooked=CTFG-PROGR  
 ‘The *calugi* (a kind of drink) is cooking.’
- b. *ahawəki iweru ridūranākre*  
 a-hawəki *iweru* Ø-r-I-d-*ura-dā*=kəre  
 2-woman calugi 3-CTFG-TRANS-3/REL-be.cooked-VERB=FUT  
 ‘Your wife will cook the *calugi*.’

#### 1.4.1.1. Causativization

Causative stems derived from unergative verbs, such as *rika* I ‘to walk’, are formed with the causativizer suffix *-dəkā* plus the verbalizer suffix *-dā*, e.g. (13). However, the causative suffix does not occur in causative stems derived from unaccusative verbs, such as *uka* II ‘to be cooked’ in (12b) above.

- (13) *habu kuladu ririradəkānāneri*  
*habu kuladu* Ø-r-I-rira-*dəkā-dā*=r-eri  
 man child 3-CTFG-TRANS-walk-CAUS-VERB=CTFG-PROGR  
 ‘The man is making the child walk.’

#### 1.4.1.2. Oblique Promotion

With a few *pseudo-transitive* verbs which take an allative or dative argument, such as *ākāraθi* II ‘to ask’, transitivity results in the promotion of the former oblique argument to direct object (examples from the Xambioá dialect):

- (14) a. *hawiki darikərekə rākārasire*  
*hawiki* dā-rikəre=kə Ø-r-Ø-ākāraθi=r-e  
 woman 3REFL-offspring=AL 3-CTFG-INTR-ask=CTFG-IMPERF  
 ‘The woman asked her son.’
- b. *hawiki darikəre ridākārasināre*  
*hawiki* dā-rikəre Ø-r-I-d-ākāraθi-dā=r-e  
 woman 3REFL-offspring 3-CTFG-TRANS-3/REL-ask-VERB=CTFG-IMPERF  
 ‘The woman questioned her son.’

#### 1.4.2. Transitive Verbs

Transitive verbs are those that take a direct object as one of their arguments. In Karajá, transitive verbs are always marked by the prefix *I-*, as shown in examples (1)-(4) above. Both transitive and intransitive valence prefixes may fuse with the preceding personal prefix under certain circumstances, such as in the second person in the centrifugal direction of the realis mood (15a). Notice that there is no fusion in the centripetal direction (15b).



- b. *haləkəe kəru rukare*  
*haləkəe kəru Ø-r-Ø-uka=r-e*  
 jaguar forehead 3-CTFG-PASS-split=CTFG-IMPERF  
 ‘The jaguar’s forehead was split.’

In the passive construction, the original O<sup>9</sup> becomes the subject, as it happens in languages such as English, for example. However, unlike English, where the agent in a passive construction can be expressed as an oblique (‘by-phrase’), in Karajá, the agent, although sometimes implicit, cannot be expressed at all. Thus, passives in Karajá are both a *backgrounding* construction, functioning to delete unknown or irrelevant subjects, and a *foregrounding* construction, since it results in the promotion of the original O to subject position (Foley and Van Valin 1985).

### 1.4.2.3. Antipassive

Antipassive is a phenomenon typical of ergative languages, corresponding functionally to a ‘mirror image’ of the passive construction in nominative-accusative languages (Silverstein 1976). In a syntactically ergative language, “while the A and the O in an ergative clause are marked as ergative and absolutive respectively, the A in an antipassive is typically coded as an absolutive NP, and the O (if present) appears in a case other than the absolutive” (Cooreman 1994:50). Although some authors, such as Cooreman, limit the discussion of antipassive constructions to ergative languages, nominative-accusative languages may also present *backgrounding* antipassives, which “function to demote the undergoer to peripheral status” (Foley and Van Valin 1985:338). This is what occurs in Karajá, where antipassive, marked by the prefix *ɔ-*, results in the deletion of an unknown or irrelevant direct object:

- (19) *nadı rəθuhɔreri*  
*d-adı Ø-r-ɔ-θuhɔ=r-eri*  
 REL-mother 3-CTFG-ANTI-wash=CTFG-PROGR  
 ‘My mother is washing (something).’

- (20) *habu rotfukareri*  
*habu Ø-r-ɔ-f-uka=r-eri*  
 man 3-CTFG-ANTI-3/REL-split=CTFG-PROGR  
 ‘The man is splitting (something).’

As these examples show, antipassive in Karajá is not promotional (or foregrounding), in the sense that the A remains in the same syntactic relation it occupies in the corresponding active, transitive voice. Furthermore, the antipas-

<sup>9</sup> I will follow Cooreman in adopting Dixon’s (1979) use of the labels A and O to refer to the two participants in a two-participant clause—prototypically, the agent and the patient, respectively.

sive construction in Karajá does not allow the expression of the demoted O whatsoever, which is an interesting parallel with what occurs to the agent in the passive construction.

### 1.5. Noun Incorporation

Noun incorporation in Karajá is a process by which the head of the absolutive noun phrase is inserted into the verb, thereby forming a compound. The more productive pattern of noun incorporation involves only body part terms, which are in general inherently possessed nouns.<sup>10</sup> Since only the head of the absolutive noun phrase is incorporated, the valence of the resulting noun-verb compound remains unaltered, as the possessor is promoted to subject with intransitive, unaccusative verbs such as *bɔhɔ* I ‘to break’, e.g. (21), or to object with transitive verbs such as *dəka* I ‘to tie’, e.g. (22).<sup>11</sup>

- (21) a. *inðwe rɪkɪ dəi rabohore*  
*idð we rɪkɪ dəi Ø-r-a-bɔhɔ=r-e*  
 people belly NARR 3.LOC 3-CTFG-INTR-break=CTFG-IMPERF  
 ‘The people’s bellies were broken there, it is said.’
- b. *inð rɪkɪ dəi rawebohore*  
*idð rɪkɪ dəi Ø-r-a-we-bɔhɔ=r-e*  
 people NARR 3.LOC 3-CTFG-INTR-belly-break=CTFG-IMPERF  
 ‘The people’s bellies were broken there, it is said.’

<sup>10</sup> Karajá also presents *classificatory noun incorporation*. In such cases, the incorporated items are body part terms that ordinarily function as *measure terms*, e.g. (a), such as *ra* ‘head’ (measure term for potatoes and yams), *ru* ‘eye’ (measure term for grains), and *we* ‘belly’ (measure term for round fruits):

- (a) *anðna i-we-θohodʒi* (b) *anðna rɪwekəkərɪ*  
*adðda ɪ-we-θohodi adðda Ø-r-i-we-kəkə=r-ɛɪ*  
 pineapple 3-belly-one pineapple 3-CTFG-TRANS-belly-grate=CTFG-PROGR  
 ‘one pineapple’ ‘He/she is grating pineapple.’

<sup>11</sup> Examples (21a) and (21b) are from the Javaé dialect. Although Maia (1998:63) claims that object incorporation does not occur in Javaé, noun incorporation seems to be as common in Javaé as it is in the other three dialects. The example below, involving the incorporation of the noun *dikɔhu* I ‘knee’ to the transitive verb *wɛ* I ‘to penetrate’, occurs in the same text from which the examples above were obtained:

- (a) *ɪdɪkɔhuwɛrɛ* *ɪdɪkɔhu rɪwɛrɛ*  
*Ø-r-i-Ø-dikɔhu-wɛ=r-e ɪ-dikɔhu Ø-r-i-wɛ=r-e*  
 3-CTFG-TRANS-3-knee-penetrate=CTFG-IMPERF 3-knee 3-CTFG-TRANS-penetrate=CTFG-IMPERF  
 ‘[He] stabbed him in the knee, he stabbed his knee.’

- (22) a. *kənãfiwe kuθehewedĩ rɪdəkare*  
*kədãfiwe kuθehewe dɪ Ø-r-ɪ-dəka=r-e*  
 K. rhea leg 3-CTFG-TRANS-tie=CTFG-IMPERF  
 ‘Kynyxiwè tied the legs of the rhea.’
- b. *kənãfiwe kuθehewe rɪdɪdəkare*  
*kədãfiwe kuθehewe Ø-r-ɪ-dɪ-dəka=r-e*  
 K. rhea 3-CTFG-TRANS-leg-tie=CTFG-IMPERF  
 ‘Kynyxiwè tied the legs of the rhea.’

As we have seen in (4) above, since noun incorporation is a valence-preserving process, an incorporating transitive verb can still be made passive or antipassive. The following section discusses the implications of the interaction between noun incorporation and antipassive in Karajá for Baker’s incorporating theory of antipassive.

## 2. Antipassive and Noun Incorporation

In this section, I will discuss the problems that the co-occurrence of antipassive and noun incorporation may potentially pose to an incorporating analysis of antipassive, taking into consideration the description of Karajá morphosyntax sketched above. Adopting the theoretical framework of Government and Binding theory, Baker (1988) proposes a treatment of familiar grammatical function changing processes such as passive, possessor raising, and applicatives as a matter of incorporation, taking incorporation in a rather broad sense, as “processes by which one semantically independent word comes to be ‘inside’ another” (1). Instead of explicit rules which would account separately for each grammatical function changing process, Baker claims that “the heart of all apparent GF changing processes is the movement of a word or (more technically) a lexical category” (19). Thus, Baker claims that words, and not only phrases, can be moved, a process he terms  $X^0$  movement, an instance of the generalized movement transformation Move-Alpha.<sup>12</sup>

Baker’s incorporation theory provides a unifying account for a number of apparently distinct grammatical function changing processes. Thus, while applicatives are seen as cases of adpositional incorporation, antipassive is treated as a matter of noun incorporation, along with noun incorporation proper and possessor raising reanalysis. According to Baker (1988:133), “antipassive is merely a special case of Noun Incorporation in which a single, designated lexical item incorporates.” The antipassive morpheme, which semantically corresponds to an indefinite NP, is generated in the direct object position at D-structure and then adjoined to the verb by  $X^0$  movement. As is typical in Government and Binding

<sup>12</sup> Given limitations of space, the summary of Baker’s theory presented here is necessarily succinct. My purpose in this section is not to provide an extensive discussion of his theory, but simply to point out its inadequacy to explain the Karajá data.

theory, the movement leaves a trace, which must be coindexed with the moved element in order to satisfy the Empty Category Principle (ECP), according to which traces must be properly governed.

In some languages, however, antipassive markers can co-occur with an incorporated noun. Karajá, as we have seen, is one of these languages. Cases such as these raise an interesting question concerning the way antipassive and noun incorporation proper would interact in an incorporating account of antipassive. Baker (1988:138) mentions the case of Nisg̃ha, cf. (23), as well as the Mayan languages as languages where antipassive marking occurs with an incorporated noun. According to him, in these languages, the antipassive morpheme “acts as a kind of ‘linking morpheme’ which appears when the object noun root is incorporated into the verb.” The antipassive “provides the theta role link necessary for Noun Incorporation to take place” (139). Baker apparently does not consider examples such as (23) as cases of double NI, since both the antipassive affix and the incorporated noun share one and the same thematic relation.

(23) Nisg̃ha (Baker 1988:138)

- |    |                                                                                |    |                                                                                                        |
|----|--------------------------------------------------------------------------------|----|--------------------------------------------------------------------------------------------------------|
| a. | <i>simiyeeni-sgu-m-hoon</i><br>smoke-APASS-ADJ- <b>fish</b><br>‘to smoke fish’ | b. | <i>lits’il-sgu-m-daala</i><br>count.up-APASS-ADJ- <b>money</b><br>‘to keep track of money (donations)’ |
|----|--------------------------------------------------------------------------------|----|--------------------------------------------------------------------------------------------------------|

All the examples cited by Baker seem to involve cases in which noun incorporation is a valence-changing process, making intransitive an otherwise transitive verb. In such cases, antipassive marking seems to be merely a consequence of the fact that the verb is made intransitive by the incorporation of its object. However, this is quite different from what happens in Karajá, where, as we have seen, productive noun incorporation is always a valence-preserving process. In cases in which the antipassive marker occurs with an incorporated noun, it is clear that the implied direct object corresponds to the original possessor of the incorporated noun (4c). If antipassive is indeed a case of noun incorporation, it is necessary to admit that, in examples such as this, antipassivization would have to be preceded by noun incorporation proper, which would place the antipassive morpheme in direct object position, making it eligible for incorporation.

In principle, this fact would apparently not be problematic for Baker’s approach, since, according to him, the antipassive does not necessarily affect only structural direct objects (Baker 1988:136). The antipassive morpheme may correspond, for example, to the subject of a lower clause, which is then moved up to a higher clause in a Raising-to-Object construction. This is what happens in Chamorro:

(24) Chamorro (Baker 1988:137)

- Kao man-ekspekta hao pära un ma’-ayuda?*  
Q APASS-expect you(ABS) IRREAL-2S-PASS-help  
‘Do you expect someone to help you?’

Thus, one might likewise postulate that in Karajá, the antipassive morpheme could be generated as a possessor, which would be promoted to direct object through the incorporation of the possessed noun. The new object would then be incorporated into the verb, through antipassivization. However, Baker categorically rules out the possibility of multiple noun incorporations:

Both case theory and the ECP rule out acyclic combinations, where first a noun incorporates into the verb and then the possessor itself incorporates—even though that possessor will look like a direct object on the surface given the GTC. (Baker 1988:374)<sup>13</sup>

According to the Government Transparency Corollary (GTC), “a lexical category which has an item incorporated into it governs everything which the incorporated item governed in its original structural position” (64). Thus, the incorporation of the former possessor should in principle be possible, since it now behaves as the direct object. However, Baker (367) argues that acyclic combinations would violate the *proper government* required by the ECP. According to this notion of government, intervening traces (such as the one left by an incorporated noun or adposition) and not only full lexical items can act as ‘possible antecedents,’ blocking the incorporation of the new direct object into the verb.

Nevertheless, such a scenario is exactly what happens in Karajá, where noun incorporation—which is essentially a possessor raising construction—clearly feeds antipassive.<sup>14</sup> Therefore, if antipassive is, in fact, a matter of noun incorporation, Karajá provides a strong counterexample to Baker’s claim that multiple noun incorporations do not occur. This apparently raises the necessity for a revision of Baker’s approach in order to allow repeated applications of noun incorporation.

### 3. Final Remarks

This paper presents a reanalysis of the Karajá verb, revealing a complex set of grammatical relation changing mechanisms, such as antipassive and noun incorporation, which were not mentioned in previous studies of the language (Fortune and Fortune 1964, Maia 1998). In particular, the interaction between antipassive and noun incorporation in Karajá has interesting implications for Baker’s (1988) incorporating account of antipassive, providing a strong counterexample to his claim against the occurrence of multiple incorporations.

---

<sup>13</sup> *Acyclic* combinations are those resulting from a movement which “reaches down more deeply into the structure than the first one does” (Baker 1988:365).

<sup>14</sup> As Baker admits, possessor raising constructions are the most likely to yield interactions between different NIs, “because by definition there are two NPs present: the possessor and the NP headed by the possessed noun.” Therefore, one would expect possessor raising to feed “Noun Incorporation proper, Antipassivization, or even Possessor Raising itself, since each of these processes is known to involve the verb and its direct object.” However, Baker considers such combinations “systematically impossible” (Baker 1988:375).

**References**

- Baker, Mark C. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.
- Cooreman, Ann. 1994. A Functional Typology of Antipassives. In Barbara Fox and Paul J. Hopper (eds.), *Voice: Form and Function*. Amsterdam: John Benjamins, 49-87.
- Dixon, Robert M. W. 1979. Ergativity. *Language* 55:59-138.
- Foley, William, and Robert Van Valin, Jr. 1985. Information Packaging in the Clause. In Timothy Shopen (ed.), *Language Typology and Syntactic Description* 1:282-364. Cambridge University Press.
- Fortune, David, and Gretchen Fortune. 1964. Karajá Grammar. Ms., Arquivo Lingüístico do Museu Nacional, Rio de Janeiro.
- Maia, Marcus. 1998. *Aspectos Tipológicos da Língua Javaé*. München: Lincom Europa.
- Ribeiro, Eduardo R. 1996. Morfologia do verbo Karajá. M.A. thesis, Universidade Federal de Goiás.
- Ribeiro, Eduardo R. 2000a. [ATR] Vowel Harmony and Palatalization in Karajá. *Santa Barbara Papers in Linguistics, Vol. 10: Proceedings of WAIL 2000*. Santa Barbara: University of California, Santa Barbara, 80-92.
- Ribeiro, Eduardo R. 2000b. Direction in Karajá. Paper presented at the VI Encuentro de Lingüística en el Noroeste. Sonora, México: Universidad de Sonora.
- Rodrigues, Aryon D. 1994. Grammatical Affinity among Tupí, Karíb, and Macro-Jê. Ms., Universidade de Brasília.
- Silverstein, Michael. 1976. Hierarchy of Features and Ergativity. In R. M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*. Canberra: Australian Institute of Aboriginal Studies.
- Wiesemann, Ursula. 1986. The Pronoun Systems of Some Jê and Macro-Jê Languages. In U. Wiesemann (ed.), *Pronominal Systems*. Tübingen: Gunther Narr Verlag, 359-80.

Department of Linguistics  
University of Chicago  
1010 E. 59<sup>th</sup> Street  
Chicago, IL 60637

erribeir@midway.uchicago.edu



# Highly-articulated Representations in Early Grammars: Evidence from English and French\*

YVAN ROSE

*Memorial University of Newfoundland*

## 0. Introduction

Since the advent of Optimality Theory, we have observed two trends in phonological theory, namely a de-emphasis on the structural relations that hold within and across segments, as well as phonetic evidence being used to explain phonological alternations. In this paper, without denying that phonetic factors are central to some aspects of phonological patterning, I argue, based on consonant harmony and metathesis data from English- and French-learning children, that the most explanatory approach to phonology is one based primarily on highly-articulated representations and headedness in constituent structure.

In brief, consonant harmony involves a featural agreement relation at a distance between two consonants. For example, in (1a), the harmonized output for an input like *duck* will surface as [gʌk] or as [dʌt], depending on which feature neutralizes the other in the child's phonology. Metathesis also involves a relation at a distance between consonants, but without feature neutralization. Metathesis instead yields a reversal in the ordering of the features found in the target word. Taking again the input *duck*, its metathesized output will be pronounced [gʌt] by the child, as in (1b).

(1) Consonant feature harmony versus metathesis: an example

- |                            |                            |
|----------------------------|----------------------------|
| a) Consonant harmony:      | b) Metathesis:             |
| • Input: <i>duck</i> [dʌk] | • Input: <i>duck</i> [dʌk] |
| • Output: [gʌk] / [dʌt]    | • Output: [gʌt]            |

A unified account of patterns such as these found in data on developmental English and French will be proposed, which builds on the differences that exist between these two languages at the level of their respective prosodic organization. As we will see, these differences have consequences on how licensing relationships between consonantal place features and heads of prosodic constituents (foot, prosodic word) take place in these target languages.

---

\* I am indebted to Heather Goad for her support and feedback throughout the course of this research, which is discussed at length in my dissertation. I am also grateful to Glyne Piggott for useful comments on my proposal. Thanks also to the participants of *BLS 27* for their comments and suggestions, especially to Gunnar Hansson, Larry Hyman, Henning Reetz, and Elizabeth Selkirk. Of course, all errors or omissions are mine. This research was supported by SSHRCC doctoral (#752-95-1415) and postdoctoral (#756-00-0235) fellowships.

The paper is organized as follows. In section 1, I introduce the consonant harmony and metathesis patterns observed in the outputs of English-learning Amahl (Smith 1973) and French-learning Clara (Rose 2000). In section 2, I outline the relevant theoretical background and assumptions on which the analysis, detailed in section 3, is based. The account builds on Goad's (2001) suggestion that licensing might play a role in consonant harmony. I propose an analysis of consonant harmony based on the constraint LICENSE, and extend this analysis to cases of place feature metathesis. A brief conclusion is offered in section 4.

## 1. Data

In this section, I introduce the relevant consonant harmony and metathesis data from Amahl and Clara. In the interest of space and clarity, I will focus only on the interaction between Dorsal and Coronal in the children's grammars.<sup>1</sup>

Starting with Amahl, we can see, in (2a), that Coronal undergoes Dorsal assimilation in CVC and CVCV [Cor...Dor] words. Conversely, as can be seen in (2b), no assimilation is found in [Dor...Cor] words.

### (2) Amahl's consonant harmony patterns (Smith 1973)

- a) [Cor...Dor]: Dorsal harmony *duck* [dʌk] → [ɟʌk]  
*chocolate* [tʃɔklɪt] → [ɟɔki:]
- b) [Dor...Cor]: No harmony *get* [get] → [ɟet]  
*greedy* [ˈɡri:di:] → [ɟi:di:]

Clara's outputs differ from Amahl's in two respects. The first regards the trigger and target features. As can be seen in (3a), in Clara's outputs, it is Dorsal which assimilates to Coronal in [Dor...Cor] CVCV words.

### (3) Clara's CVCV words (Rose 2000)

- a) [Dor...Cor]: Coronal harmony *gâteau* [gato] → [tæ'to] 'cake'  
*couleur* [kulœʁ] → [tʊ'lœʁ] 'color'
- b) [Cor...Dor]: No target inputs of this shape.<sup>a</sup>

a. This accidental gap is presumably an artifact of the relative rarity of target words of this shape in French.

The second difference between Amahl and Clara concerns the shape of the words where consonant harmony is attested. As can be seen by comparing (3) with (4), consonant harmony is observed in Clara's CVCV words only. In CVC [Dor...Cor] words in (4a), no harmony is found, even though it would be expected in light of the data in (3a). Finally, in CVC [Cor...Dor] words, in (4b), we find a pattern of place metathesis between the two input place features.

### (4) Clara's CVC words (Rose 2000)

- a) [Dor...Cor]: No harmony *goutte* [gʊt] → [gʊt] '(a) drop'  
→ [gʊtʰ]
- b) [Cor...Dor]: Metathesis *sac* [sak] → [katʃ] 'bag'  
*tigre* [tsɪg] → [kɪn] 'tiger'

<sup>1</sup> See Rose (2000, chapter 4) for a more complete analysis which includes the feature Labial as well as a comparison with another English-learning child, Trevor (Pater 1996, 1997).

In short, by comparing Amahl's and Clara's outputs, we find interesting contrasts. First, trigger and target features are predictable on a child-specific basis only. Second, the domain of application of consonant harmony varies across languages. While consonant harmony is found in Amahl's CVCV and CVC words, it is found in Clara's CVCV words only. Finally, place metathesis is observed in Clara's [Cor...Dor] CVC words.

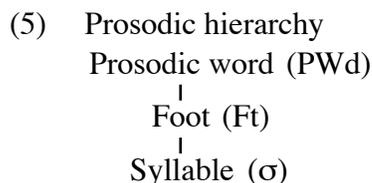
Based on these differences, I will demonstrate that a unified account of consonant harmony and metathesis must refer to highly-articulated prosodic representations at both the level of the foot and the level of the prosodic word.

## **2. Theoretical framework**

In this section, I outline the theoretical background and assumptions necessary before a satisfactory account of the patterns introduced above can be proposed. As already mentioned, much importance will be attributed to prosodic representations, which are at the core of the arguments proposed below. All aspects of the representations to be discussed are assumed to be provided by Universal Grammar (UG) as part of the child's innate linguistic competence. In order to regulate both the mapping between inputs and outputs and the licensing relations allowed in surface forms, I appeal to constraints on phonological representations, cast within Optimality Theory (OT; e.g. Prince and Smolensky 1993), which are assumed to be part of the UG endowment as well (Gnanadesikan 1995).

### **2.1. Prosodic representations**

In the analysis provided below, I integrate a set of prosodic representations, taking as a starting point the view that constituent structure is organized into the prosodic hierarchy in (5), after Selkirk (1980a,b) and McCarthy and Prince (1986).



I support the view that head-dependency relationships hold at every level of constituent structure: every branching constituent must have a head and a dependent. In addition, I assume that the relationships which take place within any category of the prosodic hierarchy in (5) are subject to the Locality condition, defined in (6), which follows the spirit of, for example, Itô (1986), Kaye (1990), and Kaye, Lowenstamm, and Vergnaud (1990).

- (6) Locality condition  
A relation is bound within the domain delimited by the highest category to which it refers.

According to the definition in (6), any relationship referring to a given prosodic domain (e.g. the foot) cannot extend beyond that domain.

The way that segmental information takes place in the hierarchy in (5) is central to the analysis detailed below. Positing the structural difference between

English and French which is relevant to the analysis requires a combination of assumptions concerning (a) the syllabification of word-final consonants, and (b) its implication for higher levels of prosodic organization in English and French. I will address these issues in turn.

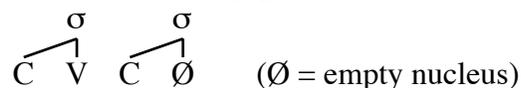
Many scholars analyze word-final consonants in the same fashion as consonants which must be syllabified word-internally outside the onset constituent, that is, as rhymal dependents (codas). This position, however, is controversial. For example, the tenets of Government Phonology (e.g. Kaye 1990, Kaye, Lowenstamm, and Vergnaud 1990, Charette 1991, Harris, 1994, 1997) hold that word-final consonants should *always* be syllabified as onsets.

Adopting a less radical position, Piggott (1999) argues that word-final consonants can be syllabified in two ways across languages: as onsets, or as codas. Piggott demonstrates that languages such as Selayarese (Mithun and Basri 1986) display a distribution of word-final consonants which matches that of word-internal codas, in the sense that they cannot license place features (they are restricted to glottal stop and placeless nasal consonants). Word-final consonants in this language are thus argued to be real codas. Piggott further argues that languages such as Diola Fogy (Sapir 1965) have a distribution of word-final consonants which is more diversified than that of word-internal codas. On the one hand, word-internal codas in Diola Fogy are restricted to the first halves of geminate nasals and sonorants that are homorganic with the following onset. From this distribution, we can infer that, similar to Selayarese, the codas of Diola Fogy cannot license place features. On the other hand, in contrast to what is observed in Selayarese, consonants with any place specification (Labial, Coronal, Dorsal) can surface word-finally in Diola Fogy. From this behavior, Piggott argues that Diola Fogy's consonants are syllabified word-finally as onsets.

Because two syllabification options are available across languages, one of these options must be universally unmarked, that is, the option that will be first entertained by a child acquiring a language with word-final consonants. Briefly addressing this issue, Piggott (1999:180) suggests that the syllabification of word-final consonants as onsets represents the unmarked case.

In the field of child language, Goad and Brannen (2000) demonstrate that word-final consonants pattern according to Piggott's suggestion: the child initially syllabifies word-final consonants as onsets, independently of the syllabification constraints of the target language. This is illustrated in (7).<sup>2</sup>

(7) CVC word in early grammars



The relationship between the syllabification in (7) and higher prosodic structure (foot, prosodic word) is at the core of the analysis proposed below. Concerning foot structure, based on the stress differences observed between English and French nouns, I assume that while the English foot is left-headed, the French foot is right-headed, as illustrated in (8).<sup>3</sup>

<sup>2</sup> Rose (2000) argues that this hypothesis holds only in contexts where word-final consonants bear place specifications, which is the case for the contexts discussed in this paper.

<sup>3</sup> Throughout the paper, the head of a given branching constituent will be represented by a vertical line linking it to its mother node.

(8) English and French foot structure

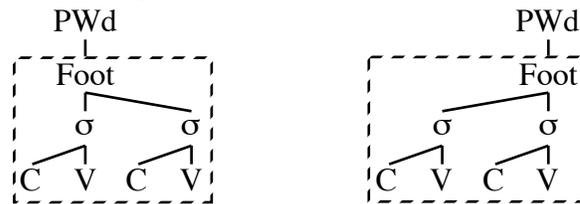
- a) English (left-headed)      b) French (right-headed)



The distinction depicted in (8) has consequences for the way that word-final consonants are linked to prosodic structure. In order to illustrate this point, I will compare the full prosodic structure of CVCV and CVC words in English and French. First, the footing of CVCV words in these languages is straightforward: all segments are dominated by the foot, as illustrated in (9).

(9) Full prosodic structure of CVCV words in English and French

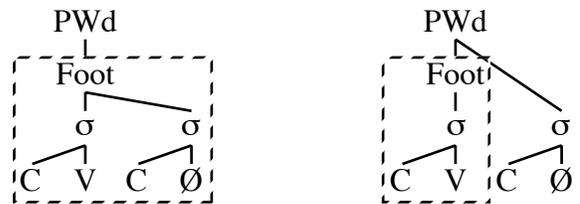
- a) CVCV in English      b) CVCV in French



The prosodification of CVC words, however, raises an interesting issue. Starting with English CVC words, I propose that these words are syllabified exactly like CVCV words and that the only distinction between English CVC and CVCV word shapes regards the (non-)realization of the final nucleus, as illustrated in (10a) (cf. (9a)). In the case of French CVC words, however, since (a) stress, which is always word-final in this language, must be realized on the only overt vowel in the word, and (b) the French foot is right-headed, I propose, following Charette (1991), that the final onset of French CVC words falls *outside* the foot, and is licensed directly by the prosodic word, as in (10b) (cf. (9b)).

(10) Full prosodic structure of CVC words in English and French

- a) CVC in English      b) CVC in French



As we will see in section 3, the claim schematized in (10) that word-final consonants are prosodified within the foot in English but outside the foot in French enables a straightforward explanation for the contrasts observed between Amahl's and Clara's place feature interaction patterns. Before I elaborate on this, it is first necessary to introduce the constraints which are relevant to the analysis.

## 2.2. Constraints

As mentioned above, in order to constrain input-output faithfulness, as well as the licensing relationships which take place between output features and heads of prosodic constituents, I assume the general framework of OT.

Concerning faithfulness relations between inputs and outputs, I appeal to correspondence constraints as proposed by McCarthy and Prince (1995). The constraints relevant to the analysis are defined in (11).

- (11) Faithfulness constraints
- a) MAX(F): Every input feature F has an output correspondent.
  - b) LINEARITY(PI, PCat): The precedence structure relative to Place specifications in the output is consistent with that of the input, and vice versa, in a given prosodic category PCat.

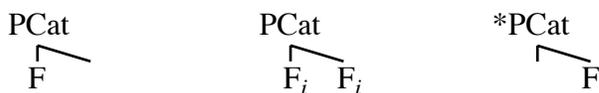
While MAX(F), in (11a), ensures preservation of input material in output forms, LINEARITY(PI, PCat), in (11b), regulates precedence Place structure between inputs and outputs, in a given prosodic category PCat.

The constraints in (11) will interact with feature licensing constraints. Assuming the Licensing Principle of Itô (1986), I propose that the aspect of licensing which is violable regards the licenser of a particular feature. In order to incorporate licensing into the general setup of OT, I appeal to the LICENSE constraint in (12), which is inspired by Itô, Mester, and Padgett (1995), Piggott (1996, 1997, 2000), and Rose (1999).

- (12) LICENSE(F, PCat): A feature F must be licensed by the head of a prosodic category PCat.

In line with Rose (1999) and Piggott (2000), I argue that LICENSE is fulfilled if and only if a segment in the *head* of PCat contains F, as schematized in (13a) and (13b). In other words, the dependent position of PCat plays no role in prosodic licensing. This implies that a feature which fails to be anchored to the head of PCat violates LICENSE, as illustrated in (13c).

- (13) LICENSE(F, PCat) relations
- a) Well-formed
  - b) Well-formed
  - c) Ill-formed



Tudanca Montañés, a dialect of European Spanish, provides us with independent evidence supporting LICENSE in adult languages. This evidence comes from a vowel centralization harmony triggered by the masculine gender suffix /-o/, which takes place between this vowel and the stressed syllable. Importantly, as demonstrated by the examples in (14), the central harmony triggered by /-o/ must (a) target the stressed vowel, but (b) never extend to the left of the stressed syllable.

- (14) Centralization in Tudeanca Montañés (Hualde 1989)<sup>a</sup>
- |               |                                |                 |
|---------------|--------------------------------|-----------------|
| [o'regæno]    | (*[o.'r <u>e</u> .gæ.no])      | ‘oregano’       |
| [anti'gwismu] | (*[an.ti.'gw <u>i</u> .si.mu]) | ‘very old (m.)’ |
| [aham'bræu]   | (*[æ.h <u>æ</u> m.'bræ.u])     | ‘hungry (m.)’   |
| [se'kælu]     | (*[s <u>e</u> .'kæ.lu])        | ‘to dry him’    |
- a. Ill-formed vowels are underlined.

To account for this system, Rose (1999) appeals to the domination of both LIC([lax], PWd) and MAX([lax]) over NOSPREAD, a general constraint against feature spreading. This analysis is summarized in the tableau in (15).<sup>4</sup>

- (15) [lax] harmony in Tudeanca Montañés (after Rose 1999)

| Input: /oreganu/ | LIC([lax], PWd) | MAX([lax]) | NOSPREAD |
|------------------|-----------------|------------|----------|
| a) [o'reganu]:   | *! ([u])        |            |          |
| b) [o'reganu]:   |                 | *!         |          |
| c) [o'regænu]:   |                 |            | **       |

In (15a), we can see that the input-like candidate incurs a fatal violation of LIC([lax], PWd): the feature [lax] of input /u/ is not licensed by the head of the prosodic word (parallel to (13c)). The two remaining candidates both satisfy LIC([lax], PWd), through feature deletion, in (15b), or feature spreading, in (15c). While (15b) fatally violates MAX([lax]), (15c), the optimal form, simultaneously satisfies the two highly-ranked constraints and only incurs violations of lower-ranked NOSPREAD.

In the next section, I analyze the child language data introduced in section 1 in a similar fashion.

### 3. Analysis

In this section, I propose an account of Amahl’s data in (2) and extend the basic approach in light of the patterns found in Clara’s CVCV and CVC words in (3) and (4), which are explained in a unified fashion. The analysis proposed, similar to the account of the Tudeanca Montañés patterns in (15), relies on the interaction between licensing and faithfulness constraints and demonstrates the effects of the structural distinctions between CVCV and CVC words in English and French discussed in section 2.1.<sup>5</sup>

#### 3.1. Amahl’s CVCV and CVC words

The constraint ranking proposed for Amahl is given in (16). This ranking is supported in (17) and (18), where Dorsal harmony in [Cor...Dor] words and the absence thereof in [Dor...Cor] words are exemplified.

<sup>4</sup> In the interest of space, no structures were included for the input forms and output candidates. To alleviate this notational limitation, feature sharing relations are represented by underlining of the segments containing the shared feature in the candidates, and the segments incurring licensing violations are indicated alongside the violation marks in the tableaux.

<sup>5</sup> While additional constraints are necessary in order to encode additional generalizations (e.g. absence of harmony between vowels and consonants; cf. Levelt 1994), for the sake of space and clarity, I will restrict the number of constraints to the ones defined in section 2.2 (see Rose 2000 for additional discussion).

- (16) Amahl’s constraint ranking  
 LIN(Pl, PwD), LIN(Pl, Ft) » MAX(Dor) » LIC(Dor, Ft) » MAX(Cor) »  
 LIC(Cor, Ft)

As we can see in (17), Dorsal harmony is correctly predicted by (16) for [Cor...Dor] input words such as *duck*. The target-like candidate in (17a), which fails to license Dorsal in the head syllable of the foot, fatally violates LIC(Dor, Ft). Place metathesis, in (17b), incurs fatal violations of the two LINEARITY constraints. Coronal harmony, in (17c), is punished because it violates MAX(Dor). Despite violating Coronal faithfulness, the Dorsal-harmonized form in (17d) satisfies the higher-ranked Dorsal faithfulness and licensing requirements of Amahl’s grammar.

- (17) Amahl’s [Cor...Dor] words

| Input:<br><i>duck</i> [dʌk] | LIN<br>(Pl, PwD) | LIN<br>(Pl, Ft) | MAX<br>(Dor) | LIC<br>(Dor, Ft) | MAX<br>(Cor) | LIC<br>(Cor, Ft) |
|-----------------------------|------------------|-----------------|--------------|------------------|--------------|------------------|
| a) [dʌk]:                   |                  |                 |              | *! ([k])         |              |                  |
| b) [gʌt]:                   | *!               | *!              |              |                  |              | * ([t])          |
| c) [dʌt]:                   |                  |                 | *!           |                  |              |                  |
| ☞ d) [gʌk]:                 |                  |                 |              |                  | *            |                  |

Turning now to [Dor...Cor] words such as *get*, in (18), we can see that the low ranking of LIC(Cor, Ft) in (16) correctly predicts disharmony: the target-like candidate in (18a) can surface as optimal.

- (18) Amahl’s [Dor...Cor] words

| Input:<br><i>get</i> [gɛt] | LIN<br>(Pl, PwD) | LIN<br>(Pl, Ft) | MAX<br>(Dor) | LIC<br>(Dor, Ft) | MAX<br>(Cor) | LIC<br>(Cor, Ft) |
|----------------------------|------------------|-----------------|--------------|------------------|--------------|------------------|
| ☞ a) [gɛt]:                |                  |                 |              |                  |              | * ([t])          |
| b) [dɛk]:                  | *!               | *!              |              | * ([k])          |              |                  |
| c) [dɛt]:                  |                  |                 | *!           |                  |              |                  |
| d) [gɛk]:                  |                  |                 |              |                  | *!           |                  |

Finally, since the domains circumscribed by the foot and the prosodic word are the same in English CVC and CVCV words (see (9a) and (10a)), the predictions made by the ranking in (16) hold for Amahl’s outputs of both shapes (cf. Clara below).

In the next subsection, I continue the exemplification of the proposal from Clara’s outputs. Recall that harmony is observed in Clara’s CVCV [Dor...Cor] words only. Both the absence of consonant harmony in CVC [Dor...Cor] words and the metathesis pattern found in CVC [Cor...Dor] words will provide additional support for the prosodic approach developed so far.

### 3.2. Clara’s CVCV words

The contrast observed between Clara’s CVCV and CVC [Dor...Cor] words ((3a) versus (4a)) is predicted by the ranking proposed in (19).



the feature Coronal can happily surface word-finally in Clara's outputs, as in the optimal form in (23d). In contrast to this, the harmonizing candidates in (23a) and (23b) both violate higher-ranked MAX requirements. Finally, (23c), which displays metathesized Dorsal in word-final position, fatally violates LIC(Dor, PWd); Dorsal fails to be realized in the head of the prosodic word (the stressed syllable) in this candidate.

(23) Clara's CVC [Dor...Cor] words<sup>a</sup>

| Input:<br><i>goutte</i> [gʊt] | LIN<br>(Pl, Ft) | MAX<br>(Cor) | MAX<br>(Dor) | LIC<br>(Dor, PWd) | LIN<br>(Pl, PWd) | LIC<br>(Cor, PWd) |
|-------------------------------|-----------------|--------------|--------------|-------------------|------------------|-------------------|
| a) [gʊk]:                     |                 | *!           |              |                   |                  |                   |
| b) [dʊt]:                     |                 |              | *!           |                   |                  |                   |
| c) [dʊk]:                     |                 |              |              | *! ([k])          | *                |                   |
| ☞ d) [gʊt]:                   |                 |              |              |                   |                  | * ([t])           |

a. Because of space limitations, the constraints LIC(Cor, Ft) and LIC(Dor, Ft), which are vacuously satisfied in French CVC words, were removed from this tableau and the following one ((24)).

The situation is different in (24), where an input [Cor...Dor] CVC word must undergo metathesis. As it was the case in (23), the two harmonizing candidates, (24a) and (24b), both fatally violate the feature faithfulness requirements of Clara's grammar. In (24c), LIC(Dor, PWd) is fatally violated by the word-final [k]. If metathesis applies, however, Coronal ends up in the word-final position, in (24d), where it surfaces without violating highly-ranked constraints.

(24) CVC [Cor...Dor] words

| Input:<br><i>sac</i> [sak] | LIN<br>(Pl, Ft) | MAX<br>(Cor) | MAX<br>(Dor) | LIC<br>(Dor, PWd) | LIN<br>(Pl, PWd) | LIC<br>(Cor, PWd) |
|----------------------------|-----------------|--------------|--------------|-------------------|------------------|-------------------|
| a) [sət]:                  |                 |              | *!           |                   |                  |                   |
| b) [xək]:                  |                 | *!           |              |                   |                  |                   |
| c) [sak]:                  |                 |              |              | *! ([k])          |                  |                   |
| ☞ d) [kas]:                |                 |              |              |                   | *                | * ([s])           |

As we can conclude from the above demonstration, the behavior of Clara's CVC words can be explained through a combination of (a) highly-articulated representations, which enable us to establish a structural distinction between non-final and final onsets with regard to how these onsets are linked to higher prosodic structure in French, and (b) a set of constraints governing feature licensing by the relevant prosodic head, similar to the analysis proposed for Tudañca Montañés. The comparison between French and English CVC words further supports this approach. The different behaviors observed in word-final consonants, which participate in consonant harmony in English but do not in French, are explained through the fact that these consonants belong to different prosodic domains in the two languages; they are prosodified within the foot in English but outside this constituent in French. Finally, while consonant harmony violates faithfulness requirements in order to satisfy higher-ranked licensing constraints, place metathesis is viewed, under the current proposal, as a strategy available to the child to ensure

that licensing requirements are satisfied, but without violating the grammar's feature faithfulness constraints.

#### **4. Conclusion**

The examples discussed in this paper cast new light concerning the structural relationships that take place between segmental features and constituent structure in early grammars. Without reference to the syllabification of word-final consonants as onsets of empty-headed syllables, and the consequence of their prosodification within the foot in English versus outside the foot in French —such a distinction can only be made by using fully-fleshed prosodic representations—, it would have been very difficult to provide a unified account of the data covered in this paper. These data therefore constitute compelling evidence in support for the view that reference to highly-articulated representations is central to the characterization of the relationships that take place within and across levels of representation in developing phonologies.

Cast in the broader context of research on acquisition, it is also important to mention that only a comparison between iambic and trochaic languages could unveil the contrasts from which the approach detailed in section 3 derives. This demonstrates that only a larger empirical base, from cross-linguistic investigations of child language data, will help us to better understand the various processes observed in language acquisition. In addition to assessing the validity of the current proposal, a comparison of the acquisition data currently available with more target languages would contribute to a better understanding of the factors governing the shapes of early grammars.

#### **References**

- Charette, Monik. 1991. *Conditions on Phonological Government*. Cambridge, MA: Cambridge University Press.
- Gnanadesikan, Amalia E. 1995. Markedness and Faithfulness Constraints in Child Phonology. Ms., University of Massachusetts, Amherst. ROA-67.
- Goad, Heather. 2001. Assimilation phenomena and initial constraint ranking in early grammars. In A.H.-J. Do, L. Dominguez, and A. Johansen, eds., *Proceedings of the 25th Annual Boston University Conference on Language Development*, 307-381. Somerville, MA: Cascadilla Press.
- Goad, Heather and Kathleen Brannen. 2000. Syllabification at the Right Edge of Words: Parallels between Child and Adult Grammars. *McGill Working Papers in Linguistics* 15:1-16
- Harris, John. 1994. *English Sound Structure*. Cambridge, MA: Blackwell.
- Harris, John. 1997. Licensing Inheritance: An Integrated Theory of Neutralisation. *Phonology* 14:315-370.
- Hualde, José I. 1989. Autosegmental and Metrical Spreading in the Vowel Harmony Systems of Northwestern Spain. *Linguistics* 27:773-805.
- Itô, Junko. 1986. Syllable Theory in Prosodic Phonology. Ph.D. diss., University of Massachusetts, Amherst. Published 1988, New York: Garland.
- Itô, Junko, Armin Mester, and Jaye Padgett. 1995. Licensing and Underspecification in Optimality Theory. *Linguistic Inquiry* 26:571-613.
- Kaye, Jonathan. 1990. 'Coda' Licensing. *Phonology* 7:301-330.

- Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1990. Constituent Structure and Government Phonology. *Phonology* 7:193-231.
- Levelt, Clara. 1994. *On the Acquisition of Place*. Leiden: Holland Institute of Generative Linguistics.
- McCarthy, John J. and Alan S. Prince. 1986. Prosodic morphology. Ms., University of Massachusetts, Amherst and Rutgers University.
- McCarthy, John J. and Alan S. Prince. 1995. Faithfulness and Reduplicative Identity. In J.N. Beckman, L. Walsh Dickey, and S. Urbanczyk, eds., *Papers in Optimality Theory*. Amherst: GLSA. 249-384.
- Mithun, Marianne and Hasan Basri. 1986. The Phonology of Selayarese. *Oceanic Linguistics* 25:210-254.
- Pater, Joseph V. 1996. Consequences of Constraint Ranking. Ph.D. diss., McGill University.
- Pater, Joseph V. 1997. Minimal Violation and Phonological Development. *Language Acquisition* 6:201-253.
- Piggott, Glyne L. 1996. Reconfiguring Harmony. *McGill Working Papers in Linguistics* 12:61-98.
- Piggott, Glyne L. 1997. Licensing and Alignment: A Conspiracy in Harmony. *Phonology* 14:437-477.
- Piggott, Glyne L. 1999. At the Right Edge of Words. *The Linguistic Review* 16:143-185.
- Piggott, Glyne L. 2000. Against Featural Alignment. *Journal of Linguistics* 36:85-129.
- Prince, Alan and Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Ms., Rutgers University and University of Colorado, Boulder.
- Rose, Yvan. 1999. Quand la structure métrique impose son accent au trait harmonique. Paper delivered at the 1999 *Annual Meeting of the Canadian Linguistic Association*. Université de Sherbrooke.
- Rose, Yvan. 2000. Headedness and Prosodic Licensing in the L1 Acquisition of Phonology. Ph.D. diss., McGill University.
- Sapir, J. David. 1965. *A Grammar of Diola-Fogny*. Cambridge: Cambridge University Press.
- Selkirk, Elisabeth O. 1980a. Prosodic Domains in Phonology: Sanskrit Revisited. In M. Aronoff and M.-L. Kean, eds., *Juncture. A Collection of Original Papers*. Saratoga, CA: Anma Libri. 107-129.
- Selkirk, Elisabeth O. 1980b. The Role of Prosodic Categories in English Word Stress. *Linguistic Inquiry* 11:563-605.
- Smith, Neilson V. 1973. *The Acquisition of Phonology, a Case Study*. Cambridge: Cambridge University Press.

Yvan Rose, Assistant Professor  
Department of Linguistics  
Memorial University of Newfoundland  
St. John's NL A1B 3X9  
CANADA

yrose@mun.ca

## On the Phonologically Driven Non-Realization of Function Words

ELISABETH SELKIRK

*University of Massachusetts, Amherst*

### 0. Introduction

The empirical focus of this paper is on cases where function words may fail to be realized in surface phonological representation due to the action of phonological constraints. Zec and Inkelas (1990) and Golston (1996) investigate cases of this sort, with a mind to their implications for theories of the organization of the grammar. My intention here is to further probe what such cases show about the nature of the phonology-morphosyntax interface, and to sketch a general theory of the phonologically driven non-realization of words.

Until recently, models of the relation between surface morphosyntax and surface phonology in a grammar did not countenance the possibility that phonology might have an influence on syntax. The idea of a phonology-free syntax was built into models of generative grammar from the 1960s up through the 1990s. These earlier models saw the output of the syntactic component as providing the input to the phonological component. But the last decade has seen possibilities for a different conception of the phonology-morphosyntax interface emerge. A Minimalist model of grammar (Chomsky 1995) does countenance the possibility that the phonological interface representation may influence syntactic form. Optimality Theory (Prince and Smolensky 1993, McCarthy 2001) allows in principle for this sort of influence as well. In the OT framework, a grammar consists of a set of ranked constraints on output form; outputs consist of both phonological structure (PStruc) and morphosyntactic structure (MStruc) representations; and constraints on PStruc and on MStruc appear in the same constraint hierarchy, such that constraints from either set may (in principle) dominate constraints from the other.

This paper adopts the OT framework in arguing for two related ways in which phonology may influence the morphosyntax of the sentence. First, the phonological constraint ranking may force the non-realization, i.e. deletion, of a word—but just when that deletion is recoverable. Second, the phonological constraint ranking may lead to the non-realization of the sentence containing the function word, i.e. to a “crashing” of the derivation, when the word deletion found in the optimal output candidate is *not* recoverable. It is proposed that both cases of non-realization arise when a particular morphosyntactic constraint of the general form

REALIZE( $\alpha$ ) (where  $\alpha$  is a variable over morphemes) is violated due to a higher ranking of phonological constraints. The question of what other types of morpho-syntactic constraints may be dominated by phonological constraints is left unaddressed in this paper.

### 1. Hausa *fa* and Phonologically Driven Non-Realization

Inkelas (1987) analyzes the possibilities of distribution of the Hausa morpheme *fa*, which she refers to as a focus particle. *Fa* is positioned by the morphosyntax to the right of the element in focus in the sentence, but it may not always appear there. For example, for a sentence to be realized in which the focus particle *fa* is associated with a preceding focused verb, that *fa* must be either verb phrase-final or be followed by a VP complement that consists of more than one word. Inkelas argues that this seemingly odd pattern of distribution has a prosodic characterization. She proposes that *fa* must appear at the right edge of a phonological phrase, as in (1b). The mechanism proposed for ensuring this distribution is a lexical prosodic subcategorization, given in (1c).

- (1) Non-realization of *fa* in Hausa (Inkelas 1987, Zec and Inkelas 1990)
- |                                                               |                                                       |
|---------------------------------------------------------------|-------------------------------------------------------|
| a. Sentence types                                             | b. Prosodic structures for those types                |
| Verb <i>fa</i>                                                | (Verb <i>fa</i> ) <sub>PPh</sub>                      |
| Verb <i>fa</i> Adj Noun                                       | (Verb <i>fa</i> ) <sub>PPh</sub> (A N) <sub>PPh</sub> |
| *Verb <i>fa</i> Noun                                          | (Verb <i>fa</i> N) <sub>PPh</sub>                     |
| c. Prosodic subcategorization frame for <i>fa</i> : [PPh ___] |                                                       |

But treating this as a case of stipulative prosodic subcategorization gives up on the search for any deeper explanation of these distributional patterns. The idea that there are no stipulative prosodic subcategorizations and that instead an output constraint hierarchy plays a determining role in the appearance, distribution, or allomorphic shape of morphemes has been proposed in Optimality Theoretic treatments of phonology-morphology interactions within words (e.g. McCarthy and Prince 1993, Mester 1994, Kager 1996). Tranel (1995), Mascaró (1996), and Golston (1996) have brought this perspective to the study of the non-realization of words in the sentence phonology. The leading idea in the theory of sentence-level allomorphy proposed by Tranel and Mascaró is that the choice between surface allomorphic variants (e.g. the appearance of French *vieux* [vjø] ‘old’ before the consonant-initial *garçon* ‘boy’ in contrast to the appearance of the allomorph *vieil* [vjɛj] ‘old’ before the vowel-initial *ami* ‘friend’) is decided by higher ranking phonological constraints, in this case the markedness constraints that favor a CV syllable structure. In other words, allomorphy is controlled by surface phonology constraints. The non-realization of morphemes in particular phonologically defined contexts, such as that of Hausa *fa*, is also arguably driven by the surface phonology.

My proposal for implementing non-realization is that a constraint of the form  $\text{REALIZE}(\alpha)$  exists for every lexical item  $\alpha$  in a language and that the language-particular ranking of  $\text{REALIZE}(\alpha)$  determines the susceptibility of  $\alpha$  to non-realization. Such a constraint in a non-morpheme-specific form has been variously termed  $\text{MORPHREAL}$ ,  $\text{EXPONENCE}$ , etc. The morphemic specificity predicts that in the grammar of Hausa, for example, there are the constraints  $\text{REALIZE}(fa)$ ,  $\text{REALIZE}(sayi)$ ,  $\text{REALIZE}(babban)$ , etc. In the default case,  $\text{REALIZE}(\alpha)$  is undominated in the constraint hierarchy, ensuring that words in the input will be realized in the output:

- (2) Default realization of morpheme  $\alpha$ :  
 $\text{REALIZE}(\alpha) \gg \text{All P-Constraints}$

In the idiosyncratic case of a particular word  $\beta$  which may fail—like *fa*—to be realized in a particular phonological context,  $\text{REALIZE}(\beta)$  is specified as lower ranked than the relevant phonological constraints on the output, giving rise to the possibility of phonologically controlled non-realization:

- (3) Idiosyncratic non-realization of morpheme  $\beta$ :  
Certain P-Constraint(s)  $\gg \text{REALIZE}(\beta)$

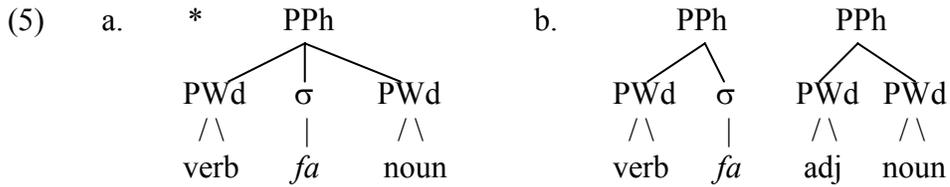
Let's see how this theory works with the case of Hausa *fa*. In the output constraint hierarchy of Hausa,  $\text{REALIZE}(fa)$  will be ranked below certain phonological constraints. It is this idiosyncratic low ranking of  $\text{REALIZE}(fa)$  which makes it vulnerable to non-realization. Because we do not have sufficient information about the prosodic structure of Hausa to make very informed hypotheses about which phonological constraints are responsible for the restrictions on the appearance of *fa*, the analysis I am about to propose can serve only as a hypothetical illustration of a type of possible analysis that could be offered, given the theory above. Let's use the cover term  $\text{PHRASING}$  for the sub-hierarchy of constraints that are responsible for the phonological phrasing of Hausa sentences that Inkelas assumes in (1b). In particular assume that a  $\text{BINARY MAXIMUM}$  constraint (cf. Selkirk 2000) will force a phonological phrase break between a verb and a two-word complement. And let's use the term  $\text{WORDING}$  as a cover term for the sub-hierarchy of constraints that are responsible for determining whether a function word is part of a Prosodic Word or not (cf. Selkirk 1995). For the sake of illustration, let's assume that  $\text{WORDING}$  in Hausa has the result that a monosyllabic function word like *fa* is not a  $\text{PWd}$  itself and is furthermore not incorporated into an adjacent  $\text{PWd}$ —which means that *fa* will be immediately dominated by a phonological phrase. And let's also posit the existence of a constraint family called  $\text{MEDIAL EXHAUSTIVITY}$ , which would rule out  $\text{EXHAUSTIVITY}$  violations except at the periphery of a prosodic constituent. This can be seen as a generalization of the peripherality constraint on “extrametricality.”

(4) MEDIAL EXHAUSTIVITY (MEDEXH):

A prosodic constituent C must immediately dominate prosodic constituents of the next level down in the prosodic hierarchy, except if the daughter constituent lies at the edge of C, e.g.

- a. \*<sub>PWd</sub>( ... Ft σ Ft ... )<sub>PWd</sub>      b. \*<sub>PPh</sub>( ... PWd σ PWd ... )<sub>PPh</sub> , etc.

Now we are in a position to account for the non-realization of *fa* in the illicit context in (5a), alongside its permissibility in (5b):



The crucial constraint ranking is that in (6):

(6) WORDING, PHRASING, MEDEXH » REALIZE(*fa*)

The success of this ranking in accounting for realization of *fa* in the (5b) context and its non-realization in the (5a) context is shown in the tableau in (7):

(7) Realization and non-realization of *fa*

| a. | [ [ [ verb ] <i>fa</i> ] [ [ adj ] [ noun ] ] ]                                                     | WORDING | PHRASING | MEDEXH | REALIZE( <i>fa</i> ) |
|----|-----------------------------------------------------------------------------------------------------|---------|----------|--------|----------------------|
| ⇒  | 1. ((verb) <sub>PWd</sub> <i>fa</i> ) <sub>PPh</sub> ((adj) <sub>PWd</sub> (noun)) <sub>PPh</sub>   |         |          |        |                      |
|    | 2. ((verb) <sub>PWd</sub> <i>fa</i> (adj) <sub>PWd</sub> (noun)) <sub>PPh</sub>                     |         | *!       | *      |                      |
|    | 3. ((verb- <i>fa</i> ) <sub>PWd</sub> ) <sub>PPh</sub> ((adj) <sub>PWd</sub> (noun)) <sub>PPh</sub> | *!      |          |        | *                    |
|    | 4. ((verb) <sub>PWd</sub> ) <sub>PPh</sub> ((adj) <sub>PWd</sub> (noun)) <sub>PPh</sub>             |         |          |        | *!                   |
| b. | [ [ [ verb ] <i>fa</i> ] [ [ noun ] ] ]                                                             | WORDING | PHRASING | MEDEXH | REALIZE( <i>fa</i> ) |
|    | 1. ( (verb) <sub>PWd</sub> <i>fa</i> ) <sub>PPh</sub> ((noun) <sub>PWd</sub> ) <sub>PPh</sub>       |         | *!       |        |                      |
|    | 2. ( (verb) <sub>PWd</sub> <i>fa</i> (noun) <sub>PWd</sub> ) <sub>PPh</sub>                         |         |          | *!     |                      |
|    | 3. ( (verb- <i>fa</i> ) <sub>PWd</sub> (noun) <sub>PWd</sub> ) <sub>PPh</sub>                       | *!      |          |        |                      |
| ⇒  | 4. ( (verb) <sub>PWd</sub> (noun) <sub>PWd</sub> ) <sub>PPh</sub>                                   |         |          |        | *                    |

In the case where the input consists of Verb-*fa* Adjective Noun, the first candidate, where *fa* lies at the edge of a phonological phrase, is the optimal one. It respects all the relevant constraints. In the case where the input consists of Verb-*fa* plus an object consisting of a single noun, it is that fourth candidate that is the optimal one. This is the one that respects all the higher constraints, but violates REALIZE(*fa*). So here the optimal candidate shows a non-realization of *fa*.

Unfortunately, Inkelas says nothing about the meaning of *fa*. Does *fa* carry some additional focus-related meaning, like ‘really’ or ‘indeed’ or ‘only’ or

‘too’? Or is it simply a semantically empty marker of a focus construction? *Fa* could in principle be semantically empty or redundant, since Focus is a property of the focused surface constituent at any rate, and is reflected in the sentence prosody (Inkelas and Leben 1990). If *fa* is empty or redundant, then its non-realization can simply be accounted for as in (7). But if *fa* does indeed have semantic content, then non-realization constitutes a violation of the principle of Recoverability:

(8) **Recoverability** (Pesetsky 1998):

A syntactic unit with semantic content must be pronounced [= realized] unless it has a sufficiently local antecedent.

So on the scenario that *fa* has semantic content (and no antecedent), if underlying *fa* were not to appear in the surface, the “derivation” of the sentence should “crash,” to use Minimalist terms. This would be a case of the non-realization of a sentence. To ensure this “derivation crashing” effect, I will assume that the following property characterizes an OT grammar:

(9) The principle of Recoverability checks the output of EVAL

If the candidate chosen by the constraint hierarchy (EVAL) involves a Recoverability violation, as would be the case in (7b) if *fa* had semantic content, then that sentence is simply not realized. This would be a case of the non-realization of a sentence, not the non-realization of a word. The theory of non-realization can be summed up as follows:

(10) **Types of Non-Realization and Recoverability:**

- i. The *non-realization of  $\alpha$*  in a phonologically illicit configuration:  
When the optimal output candidate S’ corresponding to a specific input S contains a violation of *REALIZE*( $\alpha$ ), and the absence of  $\alpha$  in the output S’ does *not* incur a Recoverability violation, then  $\alpha$  is simply not realized.
- ii. The *non-realization of a sentence with  $\alpha$*  in a phonologically illicit configuration:  
When the optimal output candidate S’ corresponding to a specific input S contains a violation of *REALIZE*( $\alpha$ ), and the absence of  $\alpha$  in the output *does* incur a Recoverability violation, then neither the optimal output S’ nor any other output candidate with the same input S is realized. (“The derivation crashes.”)

So if Hausa *fa* has no semantic content, then it will simply not be realized in the output representation corresponding to the input with *fa*. But if *fa* has semantic

content, there will be no output sentence at all corresponding to the input. It remains to see what the facts of interpretation of Hausa *fa*-less sentences are.

The above theory, then, is the alternative I am proposing to Inkelas's prosodic subcategorization theory of the non-realization of Hausa *fa*. As with allomorphic realization effects (Tranel 1995, Mascaró 1996, Golston 1996), the proposal here is that the grammar should and can shoulder the responsibility for accounting for the *non-realization* of words in phonologically defined surface configurations. What's crucial to explaining the pattern of realization of a word  $\alpha$  is (i) the lexical status of the word (what phonological shape it has, whether it is a function word or not, and whether it has semantic content), and (ii) the ranking of  $\text{REALIZE}(\alpha)$  with respect to the rest of the phonological output constraint hierarchy. This explanatory account of patterns of non-realization of words comes at a small cost—the stipulation of the ranking position of  $\text{REALIZE}(\alpha)$ .

An advantage of this theory is its ability to explain why it is that word non-realization is apparently limited to function words. A function word, unlike a “content” word, may indeed fail to make an independent semantic contribution to the sentence, and so its deletion is potentially recoverable. Second, because function words, unlike content words, may fail to be assigned the status of Prosodic Word by the constraint system (Selkirk 1995), they may violate certain phonological constraints and so be vulnerable to deletion if  $\text{REALIZE}(\alpha)$  is low ranked. The Japanese case below shows how the prosodization of the functional particle *no* determines its (non)realization.

## 2. The Recoverable Non-Realization of Japanese *no*

The case of non-realization of the Japanese particle *no* which we examine next has the advantage that the phonological constraint hierarchy that is responsible for the attested violations of  $\text{REALIZE}(no)$  has a straightforward independent motivation, and thus provides solid evidence that the non-realization of *no* is phonologically driven. The factual material is drawn from Poser (1984), who analyzes the sentence-level haplology involving adjacent instances of various functional particles with the phonological shape *-no*. These are the genitive, the copular, the nominal, and the interrogative *no*:

(11) The various *-no* particles of Japanese:

- |                 |                           |     |        |     |                          |
|-----------------|---------------------------|-----|--------|-----|--------------------------|
| a. Genitive:    | <i>Taroo no hoN</i>       |     |        |     | ‘Taro’s book’            |
|                 | Taro                      | GEN | book   |     |                          |
| b. Copular:     | <i>isya no ozisaN</i>     |     |        |     | ‘uncle, who is a doctor’ |
|                 | doctor                    | COP | uncle  |     |                          |
| c. Nominalizer: | <i>akai no</i>            |     |        |     | ‘the red one’            |
|                 | red-PRES                  | NMZ |        |     |                          |
|                 | <i>Hanako ga katta no</i> |     |        |     | ‘the one Hanako bought’  |
|                 | Hanako                    | NOM | bought | NMZ |                          |

*Phonologically-Driven Non-realization of Function Words*

- d. Interrogative: *ik-u no* 'Are you going?'  
go-PRES INT (intimate)

Given the syntax of Japanese, there are just three possible cases in which a sequence of more than one *-no* may in principle be generated in output morpho-syntactic structure. In one case, the two *-no* are realized in the output, but in the others, one of the two *-no* fails to be realized:

(12) Haplology of *-no* in Japanese (Poser 1984)

- a. *-no -no* sequence expected      output observed
- |      |         |                |
|------|---------|----------------|
| i.   | NMZ GEN | <i>-no -no</i> |
| ii.  | GEN NMZ | <i>-no</i>     |
| iii. | COP NMZ | <i>-no</i>     |

b. Examples

- |      |                                               |                               |
|------|-----------------------------------------------|-------------------------------|
| i.   | <i>akai no no futa</i>                        | ‘the red one’s lid’           |
|      | <small>red-PRES NMZ GEN lid</small>           |                               |
|      | (* <i>akai no futa</i> )                      |                               |
| ii.  | <i>ZyoN no</i> (rare)                         | ‘John’s’                      |
|      | <small>John</small>                           | [as in <i>That’s John’s</i> ] |
|      | (* <i>ZyoN no no</i> )                        |                               |
|      | <small>John GEN NMZ</small>                   |                               |
| iii. | <i>utyooten no wa Hanako de</i>               | ‘The one who is ecstatic      |
|      | <small>ecstatic TOP Hanako be</small>         | is Hanako.’                   |
|      | (* <i>utyooten no no wa Hanako de</i> )       |                               |
|      | <small>ecstatic COP NMZ TOP Hanako be</small> |                               |

What I want to show here is that the pattern of non-realization of *-no*—present in cases (ii) and (iii) and absent in (i)—is explainable in terms of an independently motivated phonological constraint hierarchy in Japanese. As for the grammaticality of outputs showing this haplologic non-realization of *-no*, I hypothesize that the genitive and copula *-no* have no semantic content, much like the particle *of* in English, and thus that their non-realization does not violate Recoverability.

The core problem to solve is why haplology appears in the contexts in (ii) and (iii) in (11), but not in the context (i). For the sake of exposition, I am going to assume a characterization of haplology in the spirit of, but not identical to, Golston (1996). According to Golston, there is a constraint he calls ANTIHOMO-PHONY that rules against adjacent segmentally identical morphemes when they are dominated by the same Prosodic Word. But a generalization of the identity conditions allows one to remove the PWD stipulation:

(13) ANTIHOMOPHONY—Generalized (cf. Golston 1996):

\*  $\alpha \beta$ , where  $\alpha$  and  $\beta$  are morphemes which are

- a) adjacent,
- b) segmentally identical, and
- c) prosodically identical

*Def.* Two morphemes of a sentence *S* are *prosodically identical* when both are dominated by identical instances of prosodic constituents in *S* and both have the same prominence status (as stressed or unstressed).

(Note that it follows from this definition that only morphemes that are dominated by the same instances of Ft, PWd, or PPh can be prosodically identical. This means that content words, which are each dominated by a distinct PWd, can never be prosodically identical, thus predicting that only function words will be susceptible to ANTIHOMOPHONY.) Haplology crucially involves a ranking of ANTIHOMOPHONY over REALIZE( $\alpha$ ):

(14) ANTIHOMOPHONY—Generalized » REALIZE( $\alpha$ )

This ranking is part of the analysis that I want to propose for *-no* haplology in Japanese. Specifically I want to propose that haplology patterns as it does in Japanese due to the manner in which the constraint hierarchy of Japanese organizes the *-no* particles into prosodic words in the different contexts in (12):

(15) The hypothesis:

- In *peripheral position* in the phonological phrase (cases (ii) and (iii)), a sequence of *-no* particles is forced to be incorporated into the same Prosodic Word with the word that precedes. ANTIHOMOPHONY rules against the sequence of non-footed *no* syllables that would have to appear within PWd, and so one of the *-no* particles fails to be realized:

\*<sub>PPh</sub>( <sub>PWd</sub>( *ZyoN no no* )<sub>PWd</sub> )<sub>PPh</sub>  
                   John     GEN NMZ

\*<sub>PPh</sub>( <sub>PWd</sub>( *utyooteN no no wa* )<sub>PWd</sub> )<sub>PPh</sub> <sub>PPh</sub>(<sub>PWd</sub>( *Hanako de* )<sub>PWd</sub> )<sub>PPh</sub>  
                   ecstatic     COP NMZ TOP                                     Hanako   be

- In *phrase-medial position* (case (i)), the first *-no* is incorporated into the preceding prosodic word, but the second one is not. ANTIHOMOPHONY does not rule the sequence out, and so haplology does not occur.

<sub>PPh</sub>( <sub>PWd</sub>( *akai no* )<sub>PWd</sub> *no* <sub>PWd</sub>( *futa* )<sub>PWd</sub> )<sub>PPh</sub>  
                   red-PRES NMZ     GEN     lid

Independent evidence for the prosodic structure analyses in (15) is provided by the phenomenon of final deaccenting (McCawley 1968 and Poser 1984).

(16) **Phrase-Final Deaccenting**

A final accent in a polysyllabic word is retained when it is followed by a phrase-final particle (*wa*, *ga*, *o*, the nominalizer *-no*, etc.), but not when the word itself is phrase-final. [Note that accent is lexically specified.]

|                                                                                           |                                                                                                                          |
|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| <p>a. <i>onna</i><br/>‘woman’<br/><i>kaki</i><br/>‘fence’<br/><i>atama</i><br/>‘head’</p> | <p>b. <i>onna’-no</i><br/>woman-NMZ ‘the woman’s’<br/><i>kaki’ wa</i><br/>fence TOP<br/><i>atama’-o</i><br/>head-ACC</p> |
|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|

The avoidance of a tonal accent on the final mora in the (a) cases can be captured by a constraint of the NONFINALITY family (Prince and Smolensky 1993), which rules out prominent syllables at edges, (17a). The ranking in (17c) of NONFINALITY(H\*, PWd) over the anti-deletion Faithfulness constraint MAX(H), (17b), provides part of the analysis of Final Deaccenting. The other part is provided by the ranking (18c) of the constraint PERIPHERAL EXHAUSTIVITY disallowing unparsed syllables at phrase edge over the constraint ALIGN-R(MWd, PWd).

(17) a. NONFINALITY(H\*, PWd):

$$*(\dots \sigma^{H*})_{PWd}$$

b. MAX(H):

A H tone in the input must be present in the output.

c. NONFINALITY(H) » MAX(H)

(18) a. PERIPHERAL EXHAUSTIVITY(PPh):

$$*(\dots \delta)_{PPh}, \text{ where } \delta \text{ is a constituent of level lower than PWd}$$

b. ALIGN-R(MWd, PWd)

Align the R edge of an MWd with the R edge of a PWd.

c. PERIPHERAL EXHAUSTIVITY » ALIGN-R(MWd, PWd)

PERIPHERAL EXHAUSTIVITY forces a phrase-final particle into the preceding PWd, where its presence saves the stem-final tone from a fatal violation of NONFINALITY:

(19) Absence of deaccenting before phrase-final particles<sup>1</sup>

| [ [onna']-ga ]                                   | PEREXH | ALIGN-R | NONFIN(H) | MAX (H) |
|--------------------------------------------------|--------|---------|-----------|---------|
| ⇒ a. ((onna'-ga) <sub>PWd</sub> ) <sub>PPH</sub> |        | *       |           |         |
| b. ((onna') <sub>PWd</sub> -ga) <sub>PPH</sub>   | *!     |         | *         |         |
| c. ((onna-ga) <sub>PWd</sub> ) <sub>PPH</sub>    |        | *       |           | *!      |
| d. ((onna) <sub>PWd</sub> -ga) <sub>PPH</sub>    | *!     |         |           | *       |

The ranking (18c), which forces phrase-final particles into the preceding PWd, will also have the effect of creating an ANTIHOMOPHONY violation in the case of a sequence of two *no*, and so leads to the haploglogic deletion of *no*:

(20) Haplogy in a phrase-final *no* sequence

| [ [utyooteN] no] no ]                                | PEREXH | ALIGN-R | ANTIOMOPH | REALIZE(α) |
|------------------------------------------------------|--------|---------|-----------|------------|
| a. ((utyooteN no no) <sub>PWd</sub> ) <sub>PPH</sub> |        | *       | *!        |            |
| b. ((utyooteN no) <sub>PWd</sub> no) <sub>PPH</sub>  | *!     | *       |           |            |
| c. ((utyooteN) <sub>PWd</sub> no no) <sub>PPH</sub>  | *!     |         |           |            |
| ⇒ d. ((utyooteN no) <sub>PWd</sub> ) <sub>PPH</sub>  |        | *       |           | *          |

So this is the story for why a sequence of two *no* is not found in phrase-final position. What now of the maintenance of the double *no* sequence in medial position within the phrase? The solution to this question also finds independent motivation in the properties of final deaccenting in phrase-medial position.

(21) **Medial Deaccenting:**

A final accent in a polysyllabic word is deleted when it is followed by a *single* phonological phrase-medial particle, e.g.

|                      |                              |
|----------------------|------------------------------|
| <i>onna no yaoya</i> | ‘the woman’s grocer’, or     |
| woman grocer         | ‘the woman, who is a grocer’ |

The analysis of this phenomenon is straightforward. It must be assumed that the *no* particle is *not* forced into a PWd with the preceding word, and that the final accent, consequently final in the PWd, therefore violates NONFINALITY and is deleted. The ranking in (22), added to the rankings motivated above, will have this result, shown in the tableau in (23):

(22) ALIGN-R(MWd, PWd) » MEDIAL EXHAUSTIVITY

<sup>1</sup> The dotted line indicates that no ranking has (as yet) been established between members of the two ranked constraint pairs posited in the tableau.

(23) Deaccenting PWd-finally when followed by single medial *-no*

| [ [onna'] no ] [yaoya]                                 | PEREXH | ALIGN-R | MEDEXH | NONFIN(H) | MAX(H) |
|--------------------------------------------------------|--------|---------|--------|-----------|--------|
| a. ((onna'-no) <sub>PWd</sub> (yaoya) <sub>PPH</sub> ) |        | *!      |        |           |        |
| b. ((onna') <sub>PWd</sub> no (yaoya)) <sub>PPH</sub>  |        |         | *      | *!        |        |
| c. ((onna-no) <sub>PWd</sub> (yaoya)) <sub>PPH</sub>   |        | *!      |        |           | *      |
| ⇒ d. ((onna) <sub>PWd</sub> no (yaoya)) <sub>PPH</sub> |        |         | *      |           | *      |

Now we are in a position to derive the failure of *no* haplology when two *no* appear in sequence phrase-medially. Our contention is that the relevant phrase in (11b,i) has the prosodic structure ((*akai-no*)<sub>PWd</sub> no (*futa*)<sub>PWd</sub>)<sub>PPH</sub> ‘the red one’s lid’, where the first *no* is incorporated into the preceding PWd, but the second is not. ANTIHOMOPHONY is not violated here, since the two *no* are not prosodically identical, not being dominated by identical constituents in the prosodic tree. The additional pairwise constraint ranking, (24), in combination with earlier rankings yields this prosodic structure as the optimal candidate in (25):

(24) ANTIHOMOPHONY » ALIGN-R(MWd, PWd)

(25) No haplology with phrase-medial sequence of *-no* particles

| [[ <i>akai</i> ] no ] no ] [ <i>futa</i> ]]                 | PEREXH | ANTIOMOPH | REALIZE(α) | ALIGN-R | MEDEXH |
|-------------------------------------------------------------|--------|-----------|------------|---------|--------|
| a. (( <i>akai no no</i> ) <sub>PWd</sub> ( <i>futa</i> ))   |        | *!        |            | *       |        |
| b. (( <i>akai</i> ) <sub>PWd</sub> no no ( <i>futa</i> ))   |        | *!        |            |         | **     |
| c. (( <i>akai no</i> ) <sub>PWd</sub> ( <i>futa</i> ))      |        |           | *!         | *       |        |
| d. (( <i>akai</i> ) <sub>PWd</sub> no ( <i>futa</i> ))      |        |           | *!         |         | *      |
| ⇒ e. (( <i>akai no</i> ) <sub>PWd</sub> no ( <i>futa</i> )) |        |           |            | *       | *      |

Safely lodged in prosodically nonidentical positions, one inside and the other outside the PWd, the two *no* in the optimal candidate are correctly predicted to surface.

This account also predicts that a final-accented word preceding a sequence of two medial *no* would fail to undergo deaccenting, in contrast to the single *no* case in (21). Unfortunately, this additional data about final deaccenting, available only from older speakers of the Tokyo standard, is not available at this writing.

### 3. Crashing Derivations vs. Alternative Syntax: Ancient Greek

Golston (1996) provides exemplary documentation from Ancient Greek to show that sequences of articles are permitted in DPs with a center-embedded possessor DP, as in (26a), but only if the adjacent articles are not homophonous, cf. (26b).

(26) a. [ *t-éei* [ *t-ées huphántikees*] *dunámei* ]  
the-D:F the-G:F weaving-G:F power-D:F  
‘with the power of weaving’

P. Pol. 281b



*Phonologically-Driven Non-realization of Function Words*

(29) Selection of (28) as optimal candidate

| [ [ <i>t-óon</i> [ <i>t-óon</i> noun ] noun ] | MORPHOSYNTAX | ANTIOMOPH | REALIZE( <i>t-óon</i> ) |
|-----------------------------------------------|--------------|-----------|-------------------------|
| a. ( <i>t-óon t-óon</i> noun)(noun)           |              | *!        |                         |
| b. ( <i>t-óon ---</i> noun)(noun)             |              |           | *!                      |
| ⇒ c. ( <i>t-óon</i> noun)( <i>t-óon</i> noun) |              |           |                         |
| d. ( <i>t-óon</i> Noun) (--- Noun)            |              |           | *!                      |

And so the impermissibility of (26b/c) would not be a case of the non-realization of a sentence. To choose between the “derivation crashing” analysis of (27) and the “alternative syntax” analysis of (29) one would need to establish whether or not these variant word orders are indeed simply alternative realizations of a same input structure with an identical semantics. In either case, it should be said, the constraint ANTIOMOPHONY and the prosodic structure constraints assigning the articles identical prosodic status are responsible for the absence of a particular syntactic construction among the output sentences of the language.

#### 4. Conclusion

To conclude, what’s been offered in this paper is a theory of the non-realization of words in specified surface phonological contexts. The claim is that the (non)recoverability of a phonologically driven deletion determines whether a word is simply deleted, or whether the sentence containing that word fails to be realized. The possibility of these limited effects of phonology on syntax are available without making any assumptions about whether syntactic constraints must in general dominate phonological constraints (Golston 1996), or whether phonological constraints may in principle dominate syntactic constraints (Zec and Inkelas 1990). The notion that Recoverability checks the output of EVAL is what allows for the phonological filtering out of syntactic constructions in these non-realization cases.

#### References

- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Golston, Chris. 1996. Syntax Outranks Phonology: Evidence from Ancient Greek. *Phonology* 12:343-368.
- Inkelas, Sharon. 1988. Prosodic Effects on Syntax: Hausa *fa*. *WCCFL* 7:375-388.
- Inkelas, Sharon, and William Leben. 1990. Where Phonology and Phonetics Intersect: The Case of Hausa Intonation. In M. Beckman and J. Kingston (eds.), *Between the Grammar and Physics of Speech*. New York: Cambridge University Press, 17-34.
- Kager, René. 1996. On Affix Allomorphy and Syllable Counting. In U. Kleinhenz (ed.), *Interfaces in Phonology*. Berlin: Akademie Verlag, 155-171.

Elisabeth Selkirk

- McCarthy, John. 2001. *Optimality Theory: A Thematic Guide*. New York: Cambridge University Press.
- McCarthy, John, and Alan Prince. 1993. *Prosodic Morphology*. Ms., University of Massachusetts and Rutgers University.
- McCawley, James. 1968. *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.
- Mascaró, Joan. 1996. External Allomorphy and Contractions in Romance. *Probus* 8:181-205.
- Pesetsky, David. 1998. Some Optimality Principles of Sentence Pronunciation. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky (eds.), *Is the Best Good Enough?* Cambridge, MA: MIT Press, 337-384.
- Poser, William. 1984. The Phonetics and Phonology of Tone and Intonation in Japanese. Ph.D. diss., MIT.
- Prince, Alan, and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Cognitive Science Center.
- Selkirk, Elisabeth. 1995. The Prosodic Structure of Function Words. In J. Beckman, L.W. Dickey, and S. Urbanczyk (eds.), *Papers in Optimality Theory: University of Massachusetts Occasional Papers* 18:349-370. Also appears in J. Morgan and K. Demuth (eds.), *Signal to Syntax*. Lawrence Erlbaum, 1996.
- Selkirk, Elisabeth. 2000. The Interaction of Constraints on Prosodic Phrasing. In M. Horne (ed.), *Prosody: Theory and Experiment*. Dordrecht: Kluwer.
- Zec, Draga, and Sharon Inkelas. 1990. Prosodically Constrained Syntax. In S. Inkelas and D. Zec (eds.), *The Phonology-Syntax Connection*. Chicago: University of Chicago Press.

Department of Linguistics  
South College  
150 Hicks Way  
University of Massachusetts  
Amherst, MA 01003-9274

selkirk@linguist.umass.edu

# Spatial Structuring in Spoken and Signed Language

LEONARD TALMY

*University at Buffalo, State University of New York*

## 0. Introduction<sup>1</sup>

Linguistic research to date has determined many of the factors that structure the spatial schemas found across spoken languages (e.g. Gruber 1965, Fillmore 1968, Leech 1969, Clark 1973, Bennett 1975, Herskovits 1982, Jackendoff 1983, Zubin and Svorou 1984, as well as myself, Talmy 1983, 2000a, 2000b). It is now feasible to integrate these factors and to determine the comprehensive system they constitute for spatial structuring in spoken language. This system is characterized by several features. With respect to constituency, there is a relatively closed universally available inventory of fundamental spatial elements that in combination form whole schemas. There is a relatively closed set of categories that these elements appear in. And there is a relatively closed small number of particular elements in each category, hence, of spatial distinctions that each category can ever mark. With respect to synthesis, selected elements of the inventory are combined in specific arrangements to make up the whole schemas represented by closed-class spatial forms. Each such whole schema that a closed-class form represents is thus a “pre-packaged” bundling together of certain elements in a particular arrangement. Each language has in its lexicon a relatively closed set of such pre-packaged schemas (larger than that of spatial closed-class forms, due to polysemy) that a speaker must select among in depicting a spatial scene. Finally, with respect to the whole schemas themselves, these schemas can undergo a certain set of processes that extend or deform them. Such processes are perhaps part of the overall system so that a language’s relatively closed set of spatial schemas can fit more spatial scenes.

An examination of signed language<sup>2</sup> shows that its structural representation of space systematically differs from that in spoken language in the direction of what

---

<sup>1</sup> An expanded and more updated version of the present paper will appear in Talmy (in press).

<sup>2</sup> I here approach signed language from the perspective of spoken language because it is not at this point an area of my expertise. For their help with my questions on signed language, my thanks to Paul Dudis, Karen Emmorey, Samuel Hawk, Nini Hoiting, Marlon Kuntze, Scott Liddell, Stephen McCullough, Dan Slobin, Ted Suppala, Alyssa Wolf, and others, - who are not responsible for my errors and oversights.

appear to be the structural characteristics of scene parsing in visual perception. Such differences include the following: Signed language can mark finer spatial distinctions with its inventory of more structural elements, more categories, and more elements per category. It represents many more of these distinctions in any particular expression. It also represents these distinctions independently in the expression, not bundled together into pre-packaged schemas. And its spatial representations are largely iconic with visible spatial characteristics. When formal linguistic investigation of signed language began several decades ago, it was important to establish in the context of that time that signed language was in fact a full genuine language, and the way to do this, it seemed, was to show that it fit the prevailing model of language, the Chomskyan-Fodorian language module. Since then, however, evidence has been steadily accruing that signed language does diverge in various respects from spoken language. The modern response to such observations - far from once again calling into question whether signed language is a genuine language - should be to rethink what the general nature of language is. Our findings suggest that instead of some discrete whole-language module, spoken language and signed language are both based on some more limited core linguistic system that then connects with different further subsystems for the full functioning of the two different language modalities.

## **1. Fundamental Space-Structuring Elements and Categories in Spoken Language**

An initial main finding emerges from analysis of the spatial schemas expressed by closed-class (grammatical) forms across spoken languages. There is a relatively closed and universally available inventory of fundamental conceptual elements that recombine in various patterns to constitute those spatial schemas. These elements fall within a relatively closed set of categories, with a relatively closed small number of elements per category.

### **1.1. The Target of Analysis**

As background to this finding, spoken languages universally exhibit two different subsystems of meaning-bearing forms. One is the “open-class” or “lexical” subsystem, comprised of elements that are great in number and readily augmented - typically, the roots of nouns, verbs, and adjectives. The other is the “closed-class” or “grammatical” subsystem, consisting of forms that are relatively few in number and difficult to augment - including such bound forms as inflections and such free forms as prepositions and conjunctions. As argued in Talmy (2000a, ch. 1), these subsystems basically perform two different functions: open-class forms largely contribute conceptual content, while closed-class forms determine conceptual structure. Accordingly, our discussion focuses on the spatial schemas represented by closed-class forms so as to examine the concepts used by language for structuring purposes.

Across spoken languages, only a portion of the closed-class subsystem regularly represents spatial schemas. We can identify the types of closed-class forms

in this portion and group them according to their kind of schema. The types of closed-class forms with schemas for paths or sites include the following: (1) forms in construction with a nominal, such as prepositions like English *across* (as in *across the field*) or noun affixes like the Finnish illative suffix *-:n* ‘into’, as well as prepositional complexes such as English *in front of* or Japanese constructions with a “locative noun” like *ue* ‘top surface’, (as in *teeburu no ue ni* ‘table GEN top at’ = “on the table”); (2) forms in construction with a verb, such as verb satellites like English *out*, *back* and *apart* (as in *They ran out / back / apart*); (3) deictic determiners and adverbs such as English *this* and *here*; (4) indefinites, interrogatives, relatives, etc., such as English *everywhere / whither / wherever*; (5) qualifiers such as English *way* and *right* (as in *It’s way / right up there*); and (6) adverbials like English *home* (as in *She isn’t home*). Types of closed-class forms with schemas for the spatial structure of objects include the following: (1) forms modifying nominals such as markers for plexity or state of boundedness, like English *-s* for multiplexing (as in *birds*) or *-ery* for debounding (as in *shrubbery*); (2) numeral classifiers like Korean *chang* ‘planar object’; and (3) forms in construction with the verb, such as some Atsugewi Cause prefixes, like *cu-* ‘as the result of a linear object moving axially into the Figure’. Finally, sets of closed-class forms that represent a particular component of a spatial event of motion/location include the following: (1) the Atsugewi verb-prefix set that represents different Figures; (2) the Atsugewi verb-suffix set that represents different Grounds (together with Paths); (3) the Atsugewi verb-prefix set that represents different Causes; and (4) the Nez Perce verb-prefix set that represents different Manners.

## **1.2. Determining the Elements and Categories**

A particular methodology is used to determine fundamental spatial elements in language. One starts with any closed-class spatial morpheme in any language, considering the full schema that it expresses and a spatial scene that it can apply to. One then determines any factor one can change in the scene so that the morpheme no longer applies to it. Each such factor must therefore correspond to an essential element in the morpheme’s schema. To illustrate, consider the English preposition *across* and the scene it refers to in *The board lay across the road*. Let us here grant the first two elements in the *across* schema (demonstrated elsewhere): (1) a Figure object (here, the board) is spatially related to a Ground object (here, the road); and (2) the Ground is ribbonal - a plane with two roughly parallel line edges that are as long as or longer than the distance between them. The remaining elements can then be readily demonstrated by the methodology. Thus, a third element is that the Figure is linear, generally bounded at both ends. If the board were instead replaced by a planar object, say, some wall siding, one could no longer use the original *across* preposition but would have to switch to the schematic domain of another preposition, that of *over*, as in *The wall siding lay over the road*. A fourth element is that the axes of the Figure and of the Ground are roughly perpendicular. If the board were instead aligned with the road, one

could no longer use the original *across* preposition but would again have to switch to another preposition, *along*, as in *The board lay along the road*. Additionally, a fifth element of the *across* schema is that the Figure is parallel to the plane of the Ground. In the referent scene, if the board were tilted away from parallel, one would have to switch to some other locution such as *The board stuck into / out of the road*. A sixth element is that the Figure is adjacent to the plane of the Ground. If the board were lowered or raised away from adjacency, even while retaining the remaining spatial relations, one would need to switch to locutions like *The board lay (buried) in the road*. / *The board was (suspended) above the road*. A seventh element is that the Figure's length is at least as great as the Ground's width. If the board were replaced by something shorter, for example, a baguette, while leaving the remaining spatial relations intact, one would have to switch from *across* to *on*, as in *The baguette lay on the road*. An eighth element is that the Figure touches both edges of the Ground. If the board in the example retained all its preceding spatial properties but were shifted axially, one would have to switch to some locution like *One end of the board lay over one edge of the road*. Finally, a ninth element is that the axis of the Figure is horizontal (the plane of the Ground is typically, but not necessarily, horizontal). Thus, if one changes the original scene to that of a spear hanging on a wall, one can use *across* if the spear is horizontal, but not if it is vertical, as in *The spear hung across the wall*. / *The spear hung up and down on the wall*. Thus, from this single example, the methodology shows that at least the following elements figure in closed-class spatial schemas: a Figure and a Ground, a point, a line, a plane, a boundary (a point as boundary to a line, a line as boundary to a plane), parallelness, perpendicularity, horizontality, adjacency (contact), and relative magnitude.

In the procedure of systematically testing candidate factors for their relevance, the elements just listed have proved to be essential to the selected schema and hence, to be in the inventory of fundamental spatial elements. But it is equally necessary to note candidates that do not prove out, so as to know which potential spatial elements do not serve a structuring function in language. In the case of *across*, for example, one can probe whether the Figure, like the board in the referent scene, must be planar - rather than simply linear - and coplanar with the plane of the Ground. It can be seen, though, that this is not an essential element to the *across* schema, since this factor can be altered in the scene by standing the board on edge without any need to alter the preposition, as in *The board lay flat / stood on edge across the railway bed*. Thus, coplanarity is not shown by *across* to be a fundamental spatial element. However, it does prove to be so in other schemas, and so in the end must be included in the inventory. This is seen for one of the schemas represented by English *over*, as in *The tapestry hung over the wall*. Here, both the Figure and Ground must be planes and coplanar with each other. If the tapestry here were changed to something linear, say, a string of beads, it is no longer appropriate to use *over* but only something like *against*, as in *The string of beads hung \*over / against the wall*. Now, another candidate element - that the Figure must be rigid, like the board in the scene - can be tested and again found to

be inessential to the *across* schema, since a flexible linear object can be substituted for the board without any need to change the preposition, as seen in *The board / The cable lay across the railway bed*. Here, however, checking this candidate factor across numerous spatial schemas in many languages might well never yield a case in which it does figure as an essential element and so would be kept off the inventory.

This methodology affords a kind of existence proof: it can demonstrate that some element does occur in the universally available inventory of structural spatial elements since it can be seen to occur in at least one closed-class spatial schema in at least one language. The procedure is repeated numerous times across many languages to build up a sizable inventory of elements essential to spatial schemas.

The next step is to discern whether the uncovered elements comprise particular structural categories and, if so, to determine what these categories are. It can be observed that for certain sets of elements, the elements in a set are mutually incompatible - only one of them can apply at a time at some point in a schema. Such sets are here taken to be basic spatial categories. Along with their members, such categories are also part of language's fundamental conceptual structuring system for space. A representative sample of these categories is presented next.

It will be seen that these categories generally have a relatively small membership. This finding depends in part on the following methodological principles. An element proposed for the inventory should be as coarse-grained as possible - that is, no more specific than is warranted by cross-schema analysis. Correlatively, in establishing a category, care must be taken that it include only the most generic elements that have actually been determined - that is, that its membership have no finer granularity than is warranted by the element-abstraction procedure. For example, the principle of mutual incompatibility yields a spatial category of "relative orientation" between two lines or planes, a category with perhaps only two member elements (both already seen in the *across* schema): approximately parallel and approximately perpendicular. Some evidence additionally suggests an intermediary "oblique" element as a third member of the category. Thus, some English speakers may distinguish a more perpendicular sense from a more oblique sense, respectively, for the two verb satellites *out* and *off*, as in *A secondary pipe branches out / off from the main sewer line*. In any case, though, the category would have no more than these two or three members. Although finer degrees of relative orientation can be distinguished by other cognitive systems, say, in visual perception and in motor control, the conceptual structuring subsystem of language does not include anything finer than the two- or three-way distinction. The procedures of schema analysis and cross-schema comparison, together with the methodological principles of maximum granularity for elements and for category membership, can lead to a determination of the number of structurally distinguished elements ever used in language for a spatial category.

### 1.3. Sample Categories and their Member Elements

The fundamental categories of spatial structure in the closed-class subsystem of spoken language fall into three classes according to the aspect of a spatial scene they pertain to: the segmentation of the scene into individual components, the properties of an individual component, and the relations of one such component to another. In a fourth class are categories of nongeometric elements frequently found in association with spatial schemas. A sampling of categories and their member elements from each of these four classes is presented next. The examples provided here are primarily drawn from English but can be readily multiplied across a diverse range of languages (see Talmy 2000a, ch. 3).

#### 1.3.1. Categories Pertaining to Scene Segmentation

The class designated as scene segmentation may include only one category, that of “major components of a scene”, and this category may contain only three member elements: the Figure, the Ground, and a secondary Reference Object. Figure and Ground were already seen for the *across* schema. Schema comparison shows the need to recognize a third scene component, the Secondary Reference Object - in fact, two forms of it: encompassive of or external to the Figure and Ground. The English preposition *near*, as in *The lamp is near the TV* specifies the location of the Figure (the lamp) only with respect to the Ground (the TV). But localizing the Figure with the preposition *above*, as in *The lamp is above the TV*, requires knowledge not only of where the Ground object is, but also of the encompassive earth-based spatial grid, in particular, of its vertical orientation. Thus, *above* requires recognizing three components within a spatial scene, a Figure, a Ground, and a Secondary Reference Object of the encompassive type. Comparably, the schema of the *past* in *John is past the border* only relates John as Figure to the border as Ground. One could say this sentence on viewing the event through binoculars from either side of the border. But *John is beyond the border* can be said only by someone on the side of the border opposite John, hence the *beyond* schema establishes a perspective point at that location as a secondary Reference Object - in this case, of the external type.

#### 1.3.2. Categories Pertaining to an Individual Scene Component

A number of categories pertain to the characteristics of an individual spatial scene component. This is usually one of the three major components resulting from scene segmentation - the Figure, Ground, or Secondary Reference Object - but it could be others, such as the path line formed by a moving Figure. One such category is that of “dimension” with four member elements: zero dimensions for a point, one for a line, two for a plane, and three for a volume. Some English prepositions require a Ground object schematizable for only one of the four dimensional possibilities. Thus, the schema of the preposition *near* as in *near the dot* requires only that the Ground object be schematizable as a point. *Along*, as in *along the trail*, requires that the Ground object be linear. *Over* as in *a tapestry over a wall* requires a planar Ground. And *throughout*, as in *cherries throughout*

*the jello*, requires a volumetric Ground. A second category is that of “number” with perhaps four members: one, two, several, and many. Some English prepositions require a Ground comprising objects in one or another of these numbers. Thus, *near* requires a Ground consisting of just one object, *between* of two objects, *among* of several objects, and *amidst* of numerous objects, as in *The basketball lay near the boulder / between the boulders / among the boulders / amidst the cornstalks*. The category of number appears to lack any further members - that is, closed-class spatial schemas in languages around the world seem never to incorporate any other number specifications - such as ‘three’ or ‘even-numbered’ or ‘too many’. A third category is that of “motive state”, with two members: motion and stationariness. Several English prepositions mark this distinction for the Figure. Thus, in one of its senses, *at* requires a stationary Figure, as in *I stayed / \*went at the library*, while *into* requires a moving Figure, as in *I went / \*stayed into the library*. Other prepositions mark this same distinction for the Ground object (in conjunction with a moving Figure). Thus, *up to* requires a stationary Ground (here, the deer), as in *The lion ran up to the deer*, while *after* requires a moving Ground as in *The lion ran after the deer*. Apparently no spatial schemas mark such additional distinctions as motion at a fast vs. slow rate, or being located at rest vs. remaining located fixedly. A fourth category is that of “state of boundedness” with two members: bounded and unbounded. The English preposition *along* requires that the path of a moving Figure be unbounded, as shown by its compatibility with a temporal phrase in *for* but not *in*, as in *I walked along the pier for 10 minutes / \*in 20 minutes*. But the spatial locution *the length of* requires a bounded path, as in *I walked the length of the pier in 20 Minutes / \*for 10 minutes*. While some spatial schemas have the bounded element at one end of a line and the unbounded element at the other end, apparently no spatial schema marks any distinctions other than the two cited states of boundedness, such as a cline of gradually increasing boundedness along a line.

Continuing the sampling of this class, a fifth category is that of “directedness” with two members: basic and reversed. A schema can require one or the other of these elements for an encompassive Ground object, as seen for the English prepositions in *The axon grew along / against the chemical gradient*, or for the Atsugewi verb satellites for (moving) ‘downstream’ and ‘upstream’. Or it can require one of the member elements for an encompassive Secondary Reference Object (here, the line), as in *Mary is ahead of / behind John in line*. A sixth category is “type of geometry” with two members: rectilinear and radial. This category can apply to an encompassive Secondary Reference Object to yield reference frames of the two geometric types. Thus, in a subtle effect, the English verb satellite *away*, as in *The boat drifted further and further away / out from the island*, tends to suggest a rectilinear reference frame in which one might picture the boat moving rightward along a corridor or sea lane with the island on the left (as if along the x-axis of a Cartesian grid). But *out* tends to suggest a radial reference frame in which the boat is seen moving from a center point along a radius through a continuum of concentric circles. The radial member of the

geometry category can involve motion about a center, along a radius, or along a periphery. The first of these is the basis for a further category, that of “orientation of spin axis”, with two members: vertical and horizontal. The English verb satellites *around* and *over* specify motion of the Figure about a vertical or horizontal spin axis, respectively, as in *The pole spun around / toppled over* and in *I turned the pail around / over*. An eighth category is “phase of matter”, with three main members, solid, liquid, and empty space, and perhaps a fourth member, fire. Thus, among the dozen or so Atsugewi verb satellites that subdivide the semantic range of English *into* plus a Ground object, the suffix *-ik’s* specifies motion horizontally into solid matter (as chopping an ax into a tree trunk), *-ic’t* specifies motion into liquid, *-ipsnu* specifies motion into the empty space of a volumetric enclosure, and *-caw* specifies motion into a fire. The phase of matter category even figures in some English prepositions, albeit covertly. Thus, *in* can apply to a Ground object of any phase of matter, whereas *inside* can apply only to one with empty space, as seen in *The rock is in / inside the box; in / \*inside the ground; in / \*inside the puddle of water; in / \*inside the fire*. A final category in this sampled series is that of “state of consolidation” with apparently two members: compact (precisional) and diffuse (approximative). The English locative prepositions *at* and *around* distinguish these two concepts, respectively, for the area surrounding a Ground object, as in *The other hiker will be waiting for you at / around the landmark*. The same distinction is marked by the two deictic adverbs in *The hiker will be waiting for you there / thereabouts*. In addition to this sampling, some ten or so further categories pertaining to properties of an individual schema component, each category with a small number of fixed contrasts, can be readily identified.

### 1.3.3. Categories Pertaining to the Relation of One Scene Component to Another

Another class of categories pertains to the relations that one scene component can bear to another. One such category was described earlier, that of “relative orientation”, with two or three members: parallel, perpendicular, and perhaps oblique. A second such category is that of “degree of remove”, of one scene component from another. This category appears to have four or five members, two with contact between the components - coincidence and adjacency – and two or three without contact - proximal, perhaps medial, and distal remove. Some pairwise contrasts in English reveal one or another of these member elements for a Figure relating to a Ground. Thus, the locution *in the front of*, as in *The carousel is in the front of the fairground*, expresses coincidence, since the carousel as Figure is represented as being located in a *part* of the fairground as Ground. But *in front of* (without a *the*) as in *The carousel is in front of the fairground*, indicates proximality, since the carousel is now located outside the fairground and near it but not touching it. The distinction between proximal and distal can be teased out by noting that *in front of* can only represent a proximal but not a distal degree of remove, as seen in that one can say *The carousel is 20 feet in front of the fairground*, but not, *\*The*

*carousel is 20 miles in front of the fairground*, whereas *above* allows both proximal and distal degrees of remove, as seen in *The hawk is 1 foot / 1 mile above the table*. The distinction between adjacency and proximality is shown by the prepositions *on* and *over*, as in *The fly is on / over the table*. Need for a fifth category member of ‘medial degree of remove’ might come from languages with a ‘here / there / yonder’ kind of distinction in their deictic adverbs or demonstratives. A third category in this series is that of “degree of dispersion” with two members: sparse and dense. To begin with, English can represent a set of multiple Figures, say, 0-dimensional peas, as adjacent to or coincident with a 1-, 2-, or 3-dimensional Ground, say, with a knife, a tabletop, or aspic, in a way neutral to the presence or absence of dispersion, as in *There are peas on the knife; on the table; in the aspic*. But in representing dispersion as present, English can (or must) indicate its degree. Thus, a sparse degree of dispersion is indicated by the addition of the locution *here and there*, optionally together with certain preposition shifts, as in *There are peas here and there on / along the knife; on / over the table; in the aspic*. And for a dense degree of dispersion, English has the three specialized forms *all along*, *all over* and *throughout*, as seen in *There are peas all along the knife; all over the table; throughout the aspic*. A fourth category is that of “path contour” with perhaps some four members: straight, arced, circular, and meandering. Some English prepositions require one or another of these contour elements for the path of a Figure moving relative to a Ground. Thus, *across* indicates a straight path, as seen in *I drove across the plateau / \*hill*, while *over* - in its usage referring to a single path line - indicates an arced contour, as in *I drove over the hill / \*plateau*. In one of its senses, *around* indicates a roughly circular path, as in *I walked around the maypole*, and *about* indicates a meandering contour, as in *I walked about the town*. Some ten or so additional categories for relating one scene component to another, again each with its own small number of member contrasts, can be readily identified.

#### **1.3.4. Nongeometric Categories**

All the preceding elements and their categories have broadly involved geometric characteristics of spatial scenes or the objects within them - that is, they have been genuinely spatial. But a number of nongeometric elements are recurrently found in association with otherwise geometric schemas. One category of such elements is that of “force dynamics” (see Talmy 2000a, ch. 7) with two members: present and absent. Thus, geometrically, the English prepositions *on* and *against* both represent a Figure in adjacent contact with a Ground, but in addition, *on* indicates that the Figure is supported against the pull of gravity through that contact while *against* indicates that it is not, as seen in *The poster is on / \*against the wall* and *The floating helium balloon is against / \*on the wall*. A second nongeometric category is that of “accompanying cognitive/affective state”, though its extent of membership is not clear. One recurrent member, however, is the attitude toward something of its being unknown, mysterious, or risky. Perhaps in combination with elements of inaccessibility or nonvisibility, this category member is associ-

ated with the Figure's location in the otherwise spatial indications of the English preposition *beyond*, whereas it is absent from the parallel locution *on the other side of*, as in *He is beyond / on the other side of the border* (both these locutions - unlike *past* seen above - are otherwise equivalent in establishing a viewpoint location as an external Secondary Reference Object). A third nongeometric category, - in the class that relates one scene component to another - is that of "relative priority", with two members: coequal and main/ancillary. The English verb satellites *together* and *along* both indicate joint participation, as seen in *I jog together / along with him*. But *together* indicates that the Figure and the Ground are coequal partners in the activity, whereas *along* indicates that the Figure entity is ancillary to the Ground entity, who would be assumed to engage in the activity even if alone (see Talmy 2000b, ch. 3).

#### 1.4. Properties of the Inventory

By our methodology, the universally available inventory of structural spatial elements includes all elements that appear in at least one closed-class spatial schema in at least one language. These elements may indeed be equivalent in their sheer availability for use in schemas. But beyond that, they appear to differ in their frequency of occurrence across schemas and languages, ranging from very common to very rare. Accordingly, the inventory of elements - and perhaps also that of categories - may have the property of being hierarchical, with entries running from the most to the least frequent. Such a hierarchy suggests asking whether the elements in the inventory, the categories in the inventory, and the elements in each category form fully closed memberships. That is, does the hierarchy end at a sharp lower boundary or trail off indefinitely? With many schemas and languages already examined, our sampling method may have yielded all the commoner elements and categories, but as the process slows down in the discovery of the rarer forms, will it asymptotically approach some complete constituency and distinctional limit in the inventory, or will it be able to go on uncovering sporadic novel forms as they develop in the course of language change?

The latter seems likelier. Exotic elements with perhaps unique occurrence in one or a few schemas in just one language can be noted, including in English. Thus, in referring to location at the interior of a wholly or partly enclosed vehicle, the prepositions *in* and *on* distinguish whether the vehicle lacks or possesses a walkway. Thus, one is *in* a car but *on* a bus, *in* a helicopter but *on* a plane, *in* a grain car but *on* a train, and *in* a rowboat but *on* a ship. Further, Fillmore has observed that this *on* also requires that the vehicle be currently in use as transport: *The children were playing in / \*on the abandoned bus in the junkyard*. Thus, schema analysis in English reveals the element '(partly) enclosed vehicle with a walkway currently in use as transport'. This is surely one of the rarer elements in schemas around the world, perhaps unique, and its existence, along with that of various others that can be found, suggests that indefinitely many more of them can sporadically arise.

In addition to being only relatively closed at its hierarchically lower end, the inventory may include some categories whose membership seems not to settle down to a small fixed set. One such category may be that of “intrinsic parts”. Frequently encountered are the five member elements ‘front’, ‘side’, ‘back’, ‘top’, and ‘bottom’, as found in the English prepositions in *The cat lay before / beside / behind / atop / beneath the TV*. But languages like Mixtec seem to distinguish a rather different set of intrinsic parts in its spatial schemas, while Makah (Matthew Davidson, personal communication) distinguishes many more and finer parts, such as with its verb suffixes for ‘at the ankle’ and ‘at the groin’.

Apart from any fuzzy lower boundary and noncoalescing categories, there does appear to exist a graduated inventory of basic spatial elements and categories that is universally available and, in particular, is relatively closed. Bowerman (e.g. 1989) has raised the main challenge to this notion. She notes, for example, that at the same time that children acquiring English learn its *in/on* distinction, children acquiring Korean learn its distinction between *kkita* ‘put [Figure] in a snug fit with [Ground]’ and *nehta* ‘put [Figure] in a loose fit with [Ground]’ she argues that since the elements ‘snug fit’ and ‘loose fit’ are presumably rare among spatial schemas across languages, they do not come from any preset inventory, one that might plausibly be innate, but rather are learned from the open-ended semantics of the adult language. My reply is that the spatial schemas of genuinely closed-class forms in Korean may well still be built from the proposed inventory elements, and that the forms she cites are actually open verbs. Open-class semantics - whether for space or other domains - seems to involve a different cognitive subsystem, drawing from finer discriminations within a broader perceptual / conceptual sphere. The Korean verbs are perhaps learned at the same age as English space-related open-class verbs like *squeeze*. Thus, English-acquiring children probably understand that *squeeze* involves centripetal pressure from encircling or bi-/multi-laterally placed Antagonists (typically the arm(s) or hand(s)) against an Agonist that resists the pressure but yields down to some smaller compass where it blocks further pressure, and hence that one can squeeze a teddy bear, a tube of toothpaste, or a rubber ball, but not a piece of string or sheet of paper, juice or sugar or the air, a tabletop or the corner of a building. Thus, Bowerman’s challenge may be directed at the wrong target, leaving intact the proposed roughly preset inventory of basic spatial building blocks.

### **1.5. Basic Elements Assembled into Whole Schemas**

The procedure so far has been analytic, starting with the whole spatial schemas expressed by closed-class forms and abstracting from them an inventory of fundamental spatial elements. But the investigation must also include a synthetic procedure: examining the ways in which individual spatial elements are assembled to constitute whole schemas. Something of such an assembly was implicit in the initial discussion of the *across* schema. But an explicit example here can better illustrate this part of the investigation.

Consider the schema represented by the English preposition *past* as in *The ball sailed past my head at exactly 3 PM*. This schema is built out of the following fundamental spatial elements (from the indicated categories) in the indicated arrangements and relationships. There are two main scene components (members of the “major scene components” category), a Figure and a Ground (here, the ball and my head, respectively). The Figure is schematizable as a 0-dimensional point (a member element of the “dimension” category). This Figure point is moving (a member element of the “motive state” category). Hence it forms a one-dimensional line (a member of the “dimension” category”). This line constitutes the Figure’s “path”. The Ground is also schematizable as a 0-dimensional point (a member of the “dimension” category). There is a point P at a proximal remove (a member of the “degree of remove” category) from the Ground point, forming a 1-dimensional line with it (a member of the “dimension” category). This line is parallel (a member of the “relative orientation” category) to the horizontal plane (a member of the “intrinsic parts” category) of the earth-based grid (a member of the major scene components” category). The Figure’s path is perpendicular (a member of the “relative orientation” category) to this line. The Figure’s path is also parallel to the horizontal plane of the earth-based grid. If the Ground object has a front, side, and back (members of the “intrinsic parts” category), then point P is proximal to the side part. A nonboundary point (a member of the “state of boundedness” category) of the Figure’s path becomes coincident (a member of the “degree of remove” category) with point P at a certain point of time.

Note that here the Figure’s path must be specified as passing through a point proximal to the Ground because if it instead passed through the Ground point, one would switch from the preposition *past* to *into*, as in *The ball sailed into my head*, and if it instead past through some distal point, one might rather say something like *The ball sailed along some ways away from my head*. And the Figure’s path must be specified both as horizontal and as located at the side portion of the Ground because, for example here, if the ball were either falling vertically or traveling horizontally at my front, one could no longer say that it sailed past my head.

The least understood aspect of the present investigation is what well-formedness conditions, if any, may govern the legality of such combinations. As yet, no obvious principles based, say, on geometric simplicity, symmetry, consistency, or the like are seen to control the patterns in which basic elements assemble into whole schemas. On the one hand, some seemingly byzantine combinations - like the schemas seen above for *across* and *past* - occur with some regularity across languages. On the other hand, much simpler combinations seem never to occur as closed-class schemas. For example, one could imagine assembling elements into the following schema: down into a surround that is radially proximal to a center point. One could invent a preposition *apit* to represent this schema, as used in *I poured water apit my house*” to refer to my pouring water down into a nearby hole dug in the field around my house. But such schemas are not found. Similarly, a number of schematic distinctions in, for example, the domain of

rotation are regularly marked by signed languages, as seen below, and could readily be represented with the inventory elements available to spoken languages, yet they largely do not occur. It could be argued that the spoken language schemas are simply the spatial structures most often encountered in everyday activity. But that would not explain why the additional sign-language schemas - presumably also reflective of everyday experience - do not show up in spoken languages. Besides, the different sets of spatial schemas found in different spoken languages are diverse enough from each other that arguing on the basis of the determinative force of everyday experience is problematic. Something else is at work but it is not yet clear what that is.

### **1.6. Properties and Processes Applying to Whole Spatial Schemas**

It was just seen that selected elements of the inventory are combined in specific arrangements to make up the whole schemas represented by closed-class spatial forms. Each such whole schema is thus a “pre-packaged” bundling together of certain elements in a particular arrangement. Each language has in its lexicon a relatively closed set of such pre-packaged schemas - one larger than that of its spatial closed-class forms, because of polysemy. A speaker of the language must select among these schemas in depicting a spatial scene. We now observe that such schemas, though composite, have a certain unitary status in their own right, and that certain quite general properties and processes can apply to them. In particular, certain properties and processes allow a schema represented by a closed-class form to generalize to a whole family of schemas. In the case of a generalizing *property*, all the schemas of a family are of equal priority. On the other hand, a generalizing *process* acts on a schema that is somehow basic, and either extends or deforms it to yield nonbasic schemas. (see Talmy 2000a ch. 1 and 3, 2000b ch. 5). Such properties and processes are perhaps part of the overall spoken-language system so that any language’s relatively closed set of spatial closed-class forms and the schemas that they basically represent can be used to match more spatial structures in a wider range of scenes.

Looking first at generalizing properties of spatial schemas, one such property is that they exhibit a topological or topology-like neutrality to certain factors of Euclidean geometry. Thus, they are magnitude neutral, as seen in such facts as that the *across* schema can apply to a situation of any size, as in *The ant crawled across my palm / The bus drove across the country*. Further, they are largely shape-neutral, as seen by such facts as that, while the *through* schema requires that the Figure form a path with linear extent, it lets that line take any contour, as in *I zig-zagged / circled through the woods*. And they are bulk-neutral, as seen by such facts as that the *along* schema requires a linear Ground without constraint on the Ground’s radial extension, as in *The caterpillar crawled up along the filament / tree trunk*. Thus, while holding to their specific constraints, schemas can vary freely in other respects and so cover a range of spatial configurations.

Among the processes that extend schemas, one is that of “extendability from the prototype”, which can serve as an alternative interpretation for some forms of

neutrality. Thus, in the case of shape, as for the *through* schema above, this schema could alternatively be conceived as prototypically involving a strait path line for the Figure, one that can then be bent to any contour. And, in the case of bulk, as for the *along* schema above, this schema could be thought prototypically to involve a purely 1-dimensional line that then can be radially inflated. Another such process is “extendability in ungoverned dimensions”. By this process, a scene component of dimensionality N in the basic form of a schema can generally be raised in dimensionality to form a line, plane, or volume aligned in a way not conflicting with the schema’s other requirements. To illustrate, it was seen earlier under the “geometric type” category that the English verb satellite *out* has a schema involving a point Figure moving along a radius away from a center point through a continuum of concentric circles, as in *The boat sailed further and further out from the island*. This schema with the Figure idealizable as a point is the basic form. But the same satellite can be used when this Figure point is extended to form a 1-dimensional line along a radius, as in *The caravan of boats sailed further and further out from the island*. And the *out* can again be used if the Figure point were instead extended as a 1-dimensional line forming a concentric circle, as in *A circular ripple spread out from where the pebble fell into the water*. In turn, such a concentric circle could be extended to fill in the interior plane, as in *The oil spread out over the water from where it spilled*. Alternatively, the concentric circle could have been extended in the vertical dimension to form a cylinder, as in *A ring of fire spread out as an advancing wall of flames*. Or again, the circle could have been extended to form a spherical shell, as in *The balloon I blew into slowly puffed out*. And such a shell can be extended to fill in the interior volume, as in *The leavened dough slowly puffed out*. One more process in this set is “extendability across motive states”. A schema basic for one motive state and Figure geometry can in general be systematically extended to another motive state and Figure geometry. For example, a closed-class form whose most basic schema pertains to a point Figure moving to form a path can generally serve as well to represent the related schema with a stationary linear Figure in the same location as the path. Thus, probably the most basic *across* schema is actually for a moving point Figure, as in *The gopher ran across the road*. By the present process, this schema can extend to the static linear Figure schema first seen in *The board lay across the road*. All the spatial properties uncovered for that static schema hold as well for the present basic dynamic schema, which in fact is the schema in which these properties originally arise.

Among the processes that deform a schema, one is that of “stretching”, which allows a slight relaxing of one of the normal constraints. Thus, in the *across* schema, where the Ground plane is either a ribbon with a long and short axis or a square with equal axes, a static linear Figure or the path of a moving point Figure must be aligned with the short Ground axis or with one of its equal axes. Accordingly, one can say *I swam across the canal* and *I swam across the square pool* when moving from one side to the other, but one cannot say *\*I swam across the canal* when moving from one end to the other. But, by moderately stretching one

axis length relative to the other, one might just about be able to say *I swam across the pool* when moving from one end to the other of an oblong pool. Another schema deforming process is that of “feature cancellation”, in which a particular complex of elements in the basic schema is omitted. Thus, the preposition *across* can be used in *The shopping cart rolled across the boulevard and was hit by an oncoming car*, even though one feature of the schema - ‘terminal point coincides with the distal edge of the Ground ribbon’ - is canceled from the Figure’s path. Further, both this feature and the feature ‘beginning point coincides with the proximal edge of the Ground ribbon’ are canceled in *The tumbleweed rolled across the prairie for an hour*. Thus, the spoken language system includes a number of generalizing properties and processes that allow the otherwise relatively closed set of abstracted or basic schemas represented in the lexicon of any single language to be applicable to a much wider range of spatial configurations.

## **2. Spatial Structuring in Signed Language**

All the preceding findings on the linguistic structuring of space have been based on the patterns found in spoken languages. The inquiry into the fundamental concept structuring system of language leads naturally to investigating its character in another major body of linguistic realization, signed language. The value in extending the inquiry in this way would be to discover whether the spatial structuring system is the same or is different in certain respects across the two language modalities, with either discovery having major consequences for cognitive theory.

In this research extension, a problematic issue is exactly what to compare between spoken and signed language. The two language systems appear to subdivide into somewhat different sets of subsystems. Thus, heuristically, the generalized spoken language system can be thought to consist of an open-class or lexical subsystem (generally representing conceptual content); a closed-class or grammatical subsystem (generally representing conceptual structure); a gradient subsystem of “vocal dynamics” (including loudness, pitch, timbre, rate, distinctness, unit separation); and an accompanying somatic subsystem (including facial expression, gesture, and “body language”). On the other hand, by one provisional proposal, the generalized sign language system might instead divide up into the following: a subsystem of lexical forms (including noun, verb, and adjective signs); an “inflectional” subsystem (including modulations of lexical signs for person, aspect); a subsystem of size-and-shape specifiers (or SASS’s); a subsystem of so-called “classifier constructions”; a gestural subsystem (along a gradient of incorporation into the preceding subsystems); a subsystem of face, head, and torso representations; a gradient subsystem of “bodily dynamics” (including amplitude, rate, distinctness, unit separation); and an associated or overlaid somatic subsystem (including further facial expression and “body language”). In particular here, the subsystem of classifier constructions – which is apparently present in all signed languages - is a formally distinct subsystem dedicated solely to the sche-

matic structural representation of objects moving or located with respect to each other in space (see Liddell forthcoming, Emmorey in press).<sup>3</sup>

The research program of comparing the representation of spatial structure across the two language modalities ultimately requires considering the two whole systems and all their subsystems. But the initial comparison - the one adopted here - should be between those portions of each system most directly involved with the representation of spatial structure. In spoken language, this is that part of the closed-class subsystem that represents spatial structure and, in signed language, it is the subsystem of classifier constructions. Spelled out, the shared properties that make this initial comparison apt include the following. First, of course, both subsystems represent objects relating to each other in space. Second, in terms of the functional distinction between “structure” and “content” described earlier, each of the subsystems is squarely on the structural side. In fact, analogous structure-content contrasts occur. Thus, the English closed-class form *into* represents the concept of a path that begins outside and ends inside an enclosure in terms of schematic structure, in contrast with the open-class verb *enter* that represents the same concept in terms of substantive content (see Talmy 2000a, ch. 1 for this structure-content distinction). Comparably, any of the formations within a classifier expression for such an outside-to-inside path represents it in terms of its schematic structure, in contrast with the unrelated lexical verb *sign* that can be glossed as ‘enter’. Third, in each subsystem, a schematic structural form within an expression in general can be semantically elaborated by a content form that joins or replaces it within the same expression. Thus, in the English sentence *I drove it (- the motorcycle-) in (to the shed)* the parenthesized forms optionally elaborate on the otherwise schematically represented Figure and Ground. Comparably, in the ASL sentence “(SHED) (MOTORCYCLE) vehicle-move-into-enclosure”, the optionally signed forms within parentheses elaborate on the otherwise schematic Figure and Ground representations within the hyphenated classifier expression.

To illustrate the classifier system, a spatial event that English could express as *The car drove past the tree* could be expressed in ASL as follows: The signer’s dominant hand, used to represent the Figure object, here has a “3 handshape” (index and middle fingers extended forward, thumb up) to represent a land vehicle. The nondominant hand, used to represent the Ground object, here involves an upright “5 handshape” (forearm held upright with the five fingers extended upward and spread apart) to represent a tree. The dominant hand is moved horizontally across the signer’s torso and past the nondominant forearm. Further though, this basic form could be modified or augmented to represent additional particulars of the referent spatial event. Thus, the dominant hand can show additional characteristics of the path. For example, the hand could move along a curved path to indicate that the road being followed was curved, it could

---

<sup>3</sup> The “classifier” label for this subsystem - originally chosen because its constructions usually include a classifier-like handshape - can be misleading. An apter term might be the “Motion-event subsystem”.

slant upward to represent an uphill course, or both could be shown together. The dominant hand can additionally show the manner of the motion. For example, as it moves along, it could oscillate up and down to indicate a bumpy ride, or move quickly to indicate a swift pace, or both could be shown together, as well as with the preceding two path properties. And the dominant hand can show additional relationships of the Figure to the Ground. For example, it could pass nearer or farther from the nondominant hand to indicate the car's distance from the tree when passing it, it could make the approach toward the nondominant hand longer (or shorter) than the trailing portion of the path to represent the comparable relationship between the car's path and the tree, or it could show both of these together or, indeed, with all the preceding additional characteristics.

The essential finding of how signed language differs from spoken language is that it more closely parallels what appear to be the structural characteristics of scene parsing in visual perception. This difference can be observed in two venues, the universally available spatial inventory and the spatial expression. These two venues are discussed next in turn.

### **2.1. In the Inventory**

The inventory of forms for representing spatial structure available to signed language has a greater total number of fundamental elements, a greater number of categories, and generally a greater number of elements per category than the spoken language inventory. More specifically, the classifier subsystem of signed language has many of the same space-structuring categories as in the closed-class subsystem of spoken language, but it also has many categories not present there, whereas spoken language may have no categories that are absent from signed language. Comparing the membership of the corresponding categories in terms of discrete elements, the number of basic elements per category in signed language ranges from being the same as that for spoken language to being very much greater. Further, though, while the membership of some categories in signed language may well consist of discrete elements, that of others appears to be gradient. Here, any procedure of tallying some fixed number of discrete elements in a category must give way to determining the approximate fineness of distinctions that can be practicably made for that category. So while some corresponding categories across the two language modalities may otherwise be quite comparable, their memberships can be of two different types, discrete vs. analog. Altogether, then, given its greater number of categories, generally larger membership per category, and a frequently gradient type of membership, the inventory of forms for building a schematic spatial representation available to the classifier subsystem of signed language is more extensive and finer than for the closed-class subsystem of spoken language. This greater extensiveness and finer granularity of spatial distinctions seems more comparable to that of spatial parsing in visual perception.

The following are some spatial categories in common across the two language modalities, but with increasing disparity in size of membership. First, some

categories appear to be quite comparable across the two modalities. Thus, both the closed-class subsystem of spoken language and the classifier subsystem of signed language structurally segment a scene into the same three components, a Figure, a Ground, and a secondary Reference Object. Both subsystems represent the category of dimensionality with the same four members – a point, a line, a plane, and a volume. And both mark the same two degrees of boundedness: bounded and unbounded.

For certain categories, signed language has just a slightly greater membership than does spoken language. Thus, for motive state, signed language structurally represents not only moving and being located, but also remaining fixedly located - a concept that spoken languages typically represent in verbs but not in their spatial preposition-like forms.

For other spatial categories, signed language has a moderately greater membership than spoken language. In some of these categories, the membership is probably gradient, but without the capacity to represent many fine distinctions clearly. Thus, signed language can apparently mark moderately more degrees of remove than spoken language's four or five members in this category. It can also apparently distinguish moderately more path lengths than the two - short and long - that spoken language marks structurally (as in English *The bug flew right / way up there*). And while spoken language can mark at most three distinctions of relative orientation - parallel, perpendicular, and oblique - signed language can distinguish a moderately greater number, for example, in the elevation of a path's angle above the horizontal, or in the angle of the Figure's axes to that of the Ground (e.g. in the placement of a pole against a wall).

Finally, there are some categories for which signed language has an indefinitely greater membership than spoken language. Thus, while spoken language structurally distinguishes some four path contours as seen in section 2.3.3, signed language can represent perhaps indefinitely many more, including zigzags, spirals, and ricochets. And for the category "locus within referent space", spoken language can structurally distinguish perhaps at most three loci relative to the speaker's location - 'here', 'there', and 'yonder' - whereas sign language can distinguish indefinitely many more within sign space.

Apart from membership differences across common categories, signed language represents some categories not found in spoken language. One such category is the relative lengths of a Figure's path before and after encounter with the Ground. Or again, signed language can represent not only the category of "degree of dispersion" (which spoken language was seen to represent in section 2.3.3), but also the category "pattern of distribution". Thus, in representing multiple Figure objects dispersed over a planar surface, it could in addition structurally indicate that these Figure objects are linear (as with dry spaghetti over a table) and are arrayed in parallel alignment, crisscrossing, or in a jumble.

This difference in the number of structurally marked spatial category and element distinctions between spoken and signed language can be highlighted with a closer analysis of a single spatial domain, that of rotational motion. As seen

earlier, the closed-class subsystem in spoken language basically represents only one category within this domain, that of “orientation of spin axis”, and within this category distinguishes only two member elements, vertical and horizontal. These two member elements are expressed, for example, by the English verb satellites *around* and *over* as in *The pole spun around / toppled over*. ASL, by contrast, distinguishes more degrees of spin axis orientation and, in addition, marks several further categories within the domain of rotation. Thus, it represents the category of “amount of rotation” and within this category can readily distinguish, say, whether the arc of a Figure’s path is less than, exactly, more than, or many times one full circuit. These are differences that English might offer for inference only from the time signature, as in *I ran around the house for 20 seconds / in 1 minute / for 2 minutes / for hours*, while using the same single spatial form *around* for all these cases. Further, while English would continue using just *around* and *over*, ASL further represents the category of “relation of the spin axis to an object’s geometry” and marks many distinctions within this category. Thus, it can structurally mark the spin axis as being located at the center of the turning object - as well as whether this object is planar like a CD disk, linear like a propeller, or an aligned cylinder like a pencil spinning on its point. It distinguishes this from the spin axis located at the boundary of the object - as well as whether the object is linear like the “hammer” swung around in a hammer toss, a transverse plane like a swinging gate, or a parallel plane like a swung cape. And it further distinguishes these from the spin axis located at a point external to the object - as well as whether the object is point-like like the earth around the sun, or linear like a spinning hoop. Finally, ASL can structurally represent the category of “uniformity of rotation” with its two member elements, uniform and nonuniform, where English could mark this distinction only with an open-class form, like the verbs in *The hanging rope spun / twisted around*, while once again continuing with the same single structural closed-class form *around*. Thus, while spoken language structurally marks only a minimal distinction of spin axis orientation throughout all these geometrically distinct forms of rotation, signed language marks more categories as well as finer distinctions within them, and a number of these appear to be distinguished as well by visual parsing of rotational movement. Overall, the additional structural spatial distinctions represented in signed language appear to be ones also regularly abstracted out in visual scene parsing and, if this can be demonstrated, would show a closer connection of signed than of spoken language to visual perception.

## **2.2. In the Expression**

The second venue, that of any single spatial expression, exhibits further respects in which signed language differs from spoken language in the apparent direction of visual scene parsing. Several of these are outlined next.

### 2.2.1. Iconic Clustering of Elements / Categories in the Expression

The structural elements of a scene of motion are clustered together in the classifier subsystem's representation of them in signed language more as they seem to be clustered in perception. When one views a motion event, such as a car driving bumpily along a curve past a tree, it is perceptually the same single object, the car, that exhibits all of the following characteristics: it has certain object properties as a Figure, it moves, it has a manner of motion, it describes a path of a particular contour, and it relates to other surrounding objects (the Ground) in its path of motion. The Ground object or objects are perceived as separate. Correspondingly, the classifier subsystem maintains exactly this pattern of clustering. It is the same single hand, the dominant hand, that exhibits the Figure characteristics, motion, manner, path contour, and relations to a Ground object. The other hand, the nondominant, separately represents the Ground object. All spoken languages diverge to a greater or lesser extent from this visual fidelity. Thus, consider one English counterpart of the event, the sentence *The car bumped along past the tree*. Here, the subject nominal, *the car*, separately represents the Figure object. The verb *bumped* clusters together the representation of the fact of motion and the manner of motion, while its sister constituent, the satellite *along* represents the presence of a path of translational motion. The preposition *past* represents the path conformation, while its sister constituent, the nominal *the tree*, represents the Ground. It in fact remains a mystery at this point in the investigation why all spoken languages using a preposition-like constituent to indicate path always conjoin it with the Ground nominal and basically never with the Figure nominal<sup>4</sup>, even though the Figure is what executes the path, and is so represented in the classifier construction of signed language.

### 2.2.2. Iconic Representation of Elements/Categories in the Expression

The classifier subsystem of signed language appears to be iconic with visual parsing not only in its clustering of spatial elements and categories, as just seen, but largely also in its representation of them. For example, it marks one basic category opposition, that between an entity and its activity, by using an object like the hand to represent an object, and motion of the hand to represent motion of the object. More specifically, the hand or other body part represents a structural entity (such as the Figure) - with the body part's configuration representing the identity or other properties of the entity - while movements or positionings of the body part represent properties of the entity's motion, location, or orientation. For example, the hand could be held flat to represent a planar object (e.g. a sheet of paper), or curved to represent a cup-shaped object. And, as seen, any such hand-shape as Figure could be moved along a variety of trajectories that represent particular path contours. But an alternative to this arrangement could be imagined.

---

<sup>4</sup> As the only apparent exception, a "demoted Figure" (see Talmy 2000b, ch. 1) can acquire either of two "demotion particles" - e.g. English *with* and *of* - that mark whether the Figure's path had a "TO" or a "FROM" vector, as seen in *The fuel tank slowly filled with gas / drained of its gas*.

The handshape could represent the path of a Figure- e.g., a fist to represent a stationary location, the outstretched fingers held flat together to represent a straight line path, the fingers in a curved plane for a curved path, and the fingers alternately forward and back for a zigzag path. Meanwhile, the hand movement could represent the Figure's shape - e.g., the hand moving in a circle to represent a round Figure and in a straight line for a linear Figure. However, no such mapping of referents to their representations is found.<sup>5</sup> Rather, the mapping in signed language is visually iconic: it assigns the representation of a material object in a scene to a material object in a classifier complex, for example, the hand, and the representation of the movements of that object in the scene to the movements of the hand. No such iconic correspondence is found in spoken language. Thus, while material objects are prototypically expressed by nouns in English, they are instead prototypically represented by verb roots in Atsugewi (see Talmy 2000b, ch. 1). And while path configurations are prototypically represented in Spanish by verbs, this is done by prepositions and satellites in English.

Finer forms of iconicity are also found within each branch of the broad entity-activity opposition. In fact, most of the spatial categories listed in section 3.2.5 that a classifier expression can represent are largely iconic with visual parsing. Thus, an entity's form is often represented by the form of the hand(s), its size by the compass of the hand(s), and its number by the number of digits or hands extended. And, among many other categories in the list, an entity's motive state, path contour, path length, manner of motion, and rate of motion are separately represented by corresponding behaviors of the hand(s). Spoken language, again, has only a bit of comparable iconicity. As examples, path length can be iconically represented in English by the vowel length of *way*, as in *The bird flew waay / waaaay / waaaaay up there*. Path length can also be semi-iconically represented by the number of iterations, as in *The bird flew up / up up / up up up and away*. Perhaps the number of an entity can be represented in some spoken language by a closed-class reduplication. But the great majority of spoken closed-class representations show no such iconicity.

The classifier subsystem is also iconic with visual parsing in its representation of temporal progression, specifically, that of a Figure's path trajectory. For example, when an ASL classifier expression represents "The car drove past the tree", the "past" path is shown by the Figure hand progressing from the nearer side of the Ground arm to a point beside it and on to its further side, much like the path progression one would see on viewing an actual car passing a tree. By contrast, nothing in any single closed-class path morpheme in a spoken language corresponds to such a progression. Iconicity of this sort can appear in spoken language only where a complex path is treated as a sequence of subparts, each with its own morphemic representation, as in *The vacuum cleaner is down around*

---

<sup>5</sup> The size and shape specifiers (SASS's) in signed languages do permit movement of the hands to trace out an object's contours, but the hands cannot at the same time adopt a shape representing the object's path.

*behind the clothes hamper*. The classifier subsystem is further iconic with visual parsing in its extensive gradience. Many of the spatial categories listed in section 3.2.5 are largely represented in gradient form in classifier expressions. Spoken language has a bit of this, as where the vowel length of a *waaay* in English can be varied continuously. But the preponderant norm is the use of discrete spatial elements, typically incorporated into distinct morphemes. For example, insofar as they represent degree of remove, the separate forms in the series *on / next to / near / away from* represent increasing distance in quantal jumps. In the classifier subsystem, the gradient capacity of two different cognitive systems, those of visual perception and of motor control, are placed in sync, whereas the closest spoken language counterpart, the spatial portion of the closed-class subsystem, by contrast relies on the principle of discrete recombination.

### **2.2.3. A Narrow Time-Space Aperture in the Expression**

Another way that the classifier expression in signed language may be more like visual perception is that it appears to be largely limited to representing a narrow time-space aperture. The tentative principle is that a classifier complex readily represents what would appear within a narrow scope of space and time if one were to zoom in with one's scope of perception around a Figure object, but little outside that narrowed scope. Hence, a classifier expression readily represents the Figure object as to its shape or type, any manipulator or instrument immediately adjacent to the Figure, the Figure's current state of Motion (motion or locatedness), the contour or direction of a moving Figure's path, and any Manner exhibited by the Figure as it moves. However, a classifier expression can little represent related factors occurring outside the current time, such as a prior cause or a follow-up consequence. And it can little represent even concurrent factors if they lie outside the immediate spatial ambit of the Figure, factors like the ongoing causal activity of an intentional Agent or other external instrumentality. By contrast, spoken languages can largely represent such nonlocal spatio-temporal factors within a single clause. In particular, such representation occurs readily in satellite-framed languages such as English (see Talmy 2000b, ch. 1 and 3). In representing a Motion event, this type of language regularly employs the satellite constituent (e.g. the verb particle in English) to represent the Path, and the main verb to represent a "co-event". The co-event is ancillary to the main Motion event and relates to it as its precursor, enabler, cause, manner, concomitant, consequence, or the like. Satellite-framed languages can certainly use this format to represent within-aperture situations that can also be represented by a classifier complex. Thus, English can say within a single clause - and ASL can sign within a single classifier expression - a motion event in which the Figure is moved by an adjacent manipulator, as in *I pinched some moss up off the rock* and *I pulled the 5-gallon bottle of water along the counter*. The same holds for a situation in which a moving Figure exhibits a concurrent Manner, as in *The cork bobbed past the seaweed*. But English can go on to use this same one-clause format to include the representation of co-events outside the aperture. Thus, English can here include

the representation of a prior causal event, as in *I kicked the football over the goalpost* (first I kicked the ball, then it moved over the goalpost). And it can represent a subsequent event, as in *They locked the prisoner into his cell (first they put him in, then they locked it)*. Within this same single-clause format, further, English can represent an Agent's concurrent causal activity outside any direct manipulation of the Figure, as in *I walked / ran / drove / flew the memo to the home office*. And English can represent a concurrent nonagentive cause of the Figure's motion, as in *The house burned down to the ground*. But ASL can represent none of the preceding sentences within a single classifier expression. For example, it cannot represent *I ran the memo to the home office* by, say, adopting the classifier for holding a thin flat object (thumb pressed against flat fingers) with the dominant hand and placing this atop the nondominant hand while moving forward with it as it shows alternating strokes of two downward pointed fingers to indicate running (or concurrently with any other indication of running). Instead a sequence of two expressions would likely be used, for example, first one for taking a memo, then one for a person speeding along.<sup>6</sup>

Comparably, in referring to a house, to represent "It burned down" one would, for example, need first to make the lexical sign for "burn up" and then (what can be treated as) a classifier expression for a structure collapsing: the hands together in an inverted "V" with the fingers interlocked and then sharply curled down. One could not represent this in a single classifier expression, say, by writhing one's fingers about as for flames as one moves them into an interlocked position and then curling them down. Though devised, these examples nevertheless show that it is physically feasible for a signed language to represent factors related to the Figure's Motion outside its immediate space-time ambit. Accordingly, the fact that signed languages, unlike spoken languages, do avoid such representations may follow from deeper structural causes, such as a greater fidelity to the characteristics of visual perception.

However apt, though, such an account leaves some facts still needing explanation. Thus, on the one hand, it makes sense that the aperture of a classifier expression is limited temporally to the present moment - this accords with our usual understanding of visual perception. But it is not clear why the aperture is also limited spatially. Visual perception is limited spatially to a narrow scope only when attention is being focused, but is otherwise able to process a wide-scoped array. Why then should classifier expressions avoid such wide spatial scope as well? Further, sign languages *can* include representation of the Ground object within a single classifier expression (typically with the nondominant hand), even where that object is not adjacent to the Figure.

---

<sup>6</sup> The behavior here of ASL cannot be explained away on the grounds that it is simply structured like a verb-framed language, since such spoken languages typically *can* represent concurrent Manner outside a narrow aperture, in effect saying something like: "I walking / running / driving / flying carried the memo to the home office".

#### 2.2.4. Many More Elements/Categories Representable Within a Single Expression

Although the spatiotemporal aperture that can be represented within a single classifier expression may be small compared to that in a spoken-language clause, the number of distinct factors within that aperture that can be represented is enormously greater. In fact, perhaps the most striking difference between the signed and the spoken representation of space in the expression is that the classifier system in signed language permits the representation of a vastly greater number of distinct spatial categories simultaneously and independently. A spoken language like English can separately represent only up to four or five different spatial categories with closed-class forms in a single clause. As illustrated in the sentence *The bat flew way back up into its niche in the cavern*, the verb is followed in turn by: a slot for indication of path length (with three members: “zero” for ‘neutral’, *way* for ‘relatively long’, *right* for ‘relatively short’); a slot for state of return (with two members: “zero” for ‘neutral’, *back* for ‘return’); a slot for displacement within the earth-frame (with four members: “zero” for ‘neutral’, *up* for ‘positive vertical displacement’, *down* for ‘negative vertical displacement’, *over* for ‘horizontal displacement’); a slot for geometric conformation (with many members, including *in*, *across*, *past*); and perhaps a slot for motive state and vector (with two members: “zero” for ‘neutral between location AT and motion TO’ as seen in *in / on*, and *-to* for ‘motion TO’ as seen in *into / onto*). Even a polysynthetic language like Atsugewi has closed-class slots within a single clause for only up to six spatial categories: path conformation combined with Ground type, path length, vector, deixis, state of return, and cause or manner. In contrast, by one tentative count, ASL has provision for the separate indication of thirty different spatial categories. These categories do exhibit certain cooccurrence restrictions, they differ in obligatoriness or optionality, and it is unlikely - perhaps impossible - for all thirty of them to be represented at once. Nevertheless, a sizable number of them can be represented in a single classifier expression and varied independently there. The table below lists the spatial categories that I have provisionally identified as available for concurrent independent representation. The guiding principle for positing a category has been that its elements are mutually exclusive: different elements in the same category cannot be represented together in the same classifier expression. If certain elements can be concurrently represented, they belong to different categories. Following this principle has, on the one hand, involved joining together what some sign language analyses have treated as separate factors. For example, the first category below covers equally the representation of Figure, instrument, or manipulator (handling classifier), since these three kinds of elements apparently cannot be separately represented in a single expression - one or another of them must be selected. On the other hand, the principle requires making distinctions within some categories that spoken languages treat as uniform. Thus, the single “manner” category of English must be subdivided into a category of “divertive manner” (e.g. moving along with an up-down bump) and a category of “dynamic manner” (e.g. moving along rapidly)

## *Spatial Structuring in Spoken and Signed Language*

because these two factors can be represented concurrently and varied independently.

- A. entity properties
  - 1. identity (form or semantic category) of Figure / instrument / manipulator
  - 2. identity (form or semantic category) of Ground
  - 3. magnitude of some major entity dimension
  - 4. magnitude of a transverse dimension
  - 5. number of entities
- B. orientation properties
  - 6. an entity's rotatedness about its left-right axis ("pitch")
  - 7. an entity's rotatedness about its front-back axis ("roll")
  - 8. a. an entity's rotatedness about its top-bottom axis ("yaw")
    - b. an entity's rotatedness relative to its path of forward motion
- C. locus properties
  - 9. locus within sign space
- D. Motion properties
  - 10. motive state (moving / resting / fixed)
  - 11. internal motion (e.g. expansion/contraction, form change, wriggle, swirling)
  - 12. confined motion ( e.g. straight oscillation, rotary oscillation, rotation, local wander)
  - 13. translational motion
- E. Path properties
  - 14. state of continuity (unbroken / saltatory)
  - 15. contour of path
  - 16. state of boundedness (bounded / unbounded)
  - 17. length of path
  - 18. vertical height
  - 19. horizontal distance from signer
  - 20. left-right positioning
  - 21. up-down angle ("elevation")
  - 22. left-right angle ("direction")
  - 23. transitions between motion and stationariness (e.g. normal, decelerated, abrupt as from impact)
- F. Manner properties
  - 24. divertive manner
  - 25. dynamic manner
- G. relations of Figure or Path to Ground
  - 26. path's conformation relative to Ground
  - 27. relative lengths of path before and after encounter with Ground
  - 28. Figure's path relative to the Path of a moving Ground
  - 29. Figure's proximity to Ground
  - 30. Figure's orientation relative to Ground

It seems probable that something more on the order of this number of spatial categories are concurrently analyzed out by visual processing on viewing a scene than the much smaller number present in even the most extreme spoken language patterns.

### **2.2.5. Elements/Categories Independently Variable in the Expression – Not in Pre-Packaged Schemas**

The signed-spoken language difference just presented was mainly considered for the sheer number of distinct spatial categories that can be represented together in a single classifier expression. Now, though, we stress the corollary: their independent variability. That is, apart from certain constraints involving cooccurrence and obligatoriness in a classifier expression, a signer can generally select a category for inclusion independently of other categories, and select a member element within each category independently of other selections. For example, a classifier expression can separately include and independently vary a path's contour, length, vertical angle, horizontal angle, speed, accompanying manner, and relation to Ground object. By contrast, it was seen earlier that spoken languages largely bundle together a choice of spatial member elements within a selection of spatial categories for representation within the single complex schema that is associated with a closed-class morpheme. The lexicon of each spoken language will have available a certain number of such "prepackaged" spatial schemas, and the speaker must generally choose from among those to represent a spatial scene, even where the fit is not exact. The system of generalizing properties and processes seen in section 2.6 that apply to the set of basic schemas in the lexicon (including their plastic extension and deformation) may exist to compensate for the pre-packaging and closed stock of the schemas in any spoken language. Thus, what are largely semantic components within a single morpheme in spoken language correspond to what can be considered separate individually controllable morphemes in the signed classifier expression. Classifier expressions', apparent general lack of pre-packaging, of a fixed set of discrete basic schemas, or of a system for generalizing, extending, or deforming such basic schemas may well accord with comparable characteristics of visual parsing. That is, the visual processing of a viewed scene may tend toward the independent assessment of spatial factors without much pre-packeting of associated factors or of their plastic alteration. If shown to be the case, then signed language will once again prove to be closer to perceptual spatial structuring than spoken language is.

### **3. Cognitive Implications of Spoken/Signed Language Differences**

The preceding comparison of the space-structuring subsystems of spoken and of signed language has shown a number of respects in which these are similar and in which they are different. It can be theorized that their common characteristics are the product of a single neural system, what can be assumed to be the core language system, while each set of distinct characteristics results from the activity of some further distinct neural system. These ideas are outlined next.

### **3.1. Where Signed and Spoken Language are Alike**

We can first summarize and partly extend the properties above found to hold both in the closed-class subsystem of spoken language and in the classifier subsystem of signed language. Both subsystems can represent multifarious and subtly distinct spatial situations - that is, situations of objects moving or located with respect to each other in space. Both represent such spatial situations schematically and structurally. Both have basic elements that in combination make up the structural schematizations. Both group their basic elements within certain categories that themselves represent particular categories of spatial structure. Both have certain conditions on the combination of basic elements and categories into a full structural schematization. Both have conditions on the cooccurrence and sequencing of such schematizations within a larger spatial expression. Both permit semantic amplification of certain elements or parts of a schematization by open-class or lexical forms outside the schema. And in both subsystems, a spatial situation can often be conceptualized in more than one way, so that it is amenable to alternative schematizations.

### **3.2. Where Spoken and Signed Language Differ**

First, the two language modalities have been seen to divide up into somewhat different sets of subsystems without clear one-to-one matchups. Thus, the spatial portion of the spoken language closed-class subsystem and the classifier subsystem of signed language may not be exactly corresponding counterparts, but only those parts of the two language modalities closest to each other in the representation of schematic spatial structure. Within this initial comparison, though, the classifier subsystem seems closer to the structural characteristics of visual parsing than the closed-class subsystem in the following ways: It has more basic elements, categories, and elements per category in its schematic representation of spatial structure. Its elements exhibit more iconicity with the visual in the pattern in which they are clustered in an expression, in their physical representation, in their progression through time, and in their gradient character. It can represent only a narrow temporal aperture in an expression (and only a narrow spatial aperture as well, though this difference from spoken language might not reflect visual fidelity). It can represent many more distinct elements and categories together in a single expression. It can more readily select categories and category elements independently for representation in an expression. And it avoids pre-packaged category-element combinations as well as generalizations of their range and processes for their extension or deformation.

### **3.3. A New Neural Model**

In its strong reading, the Fodor-Chomsky model relevant here is of a complete inviolate language module in the brain, one that performs all and only the functions of language without influence from outside itself - a specifically linguistic "organ". But the evidence assembled here challenges such a model. What has here been found is that two different linguistic systems, the spoken and the signed,

both of them undeniably forms of human language, on the one hand share extensive similarities but - crucially - also exhibit substantial differences in structure and organization. A new neural model can be proposed that is sensitive to this finding. We can posit a “core” language system in the brain, more limited in scope than the Fodor-Chomsky module, that is responsible for the properties and performs the functions found to be in common across both the spoken and the signed modalities. In representing at least spatial structure, this core system would then further connect with two different outside brain systems responsible, respectively, for the properties and functions specific to each of the two language modalities. It would thus be the interaction of the core linguistic system with one of the outside systems that would underlie the full functioning of each of the two language modalities.

The particular properties and functions that the core language system would provide would include all the spoken-signed language properties in section 4.1 specific to spatial representation, though presumably in a more generic form. Thus, the core language system might have provision for: using individual unit concepts as the basis for representing broader conceptual content; grouping individual concepts into categories; associating individual concepts with overt physical representations, whether vocal or manual; combining individual concepts - and their physical representations - under certain constraints to represent a conceptual complex (i.e. the basis for morphosyntax); and establishing a subset of individual concepts as the basic schematic concepts that, in combinations, represent conceptual structure.

When in use for signed language, this core language system might then further connect with particular parts of the neural system for visual perception. I have previously called attention to the already great overlap of structural properties between spoken language and visual perception (see Talmy 2000a, ch. 2), which might speak to some neural connection already in place between the core language system and the visual system. Accordingly, the proposal here is that in the case of signed language, still further connections are brought into play, ones that might underlie the finer granularity, iconicity, gradience, and aperture limitations we have seen in signed spatial representations.

When in use for spoken language, the core language system might further connect with a putative neural system responsible for some of the characteristics present in spoken spatial representations but absent from signed ones. These could include the packeting of spatial elements into a stable closed set of patterned combinations, and a system for generalizing, extending, and deforming the packets. It is not clear why such a further system might otherwise exist but, very speculatively, one might look to see if any comparable operations hold, say, for the maintenance and modification of motor patterns.

The present proposal of a more limited core language system connecting with outlying subsystems for full language function seems more consonant with contemporary neuroscientific findings that relatively smaller neural assemblies link up in larger combinations in the subservance of any particular cognitive

function. In turn, the proposed core language system might itself be found to consist of an association and interaction of still smaller units of neural organization, many of which might in turn participate in subserving more than just language functions.

## References

- Bennett, David C. 1975. *Spatial and Temporal Uses of English Prepositions: An Essay in Stratificational Semantics*. London: Longman.
- Bowerman, Melissa. 1989. Learning a Semantic System: What Role Do Cognitive Predispositions Play? In Mabel L. Rice & Richard L. Schiefelbusch (eds.) *The Teachability of Language*. Baltimore: P.H. Brookes Pub. Co.
- Clark, Herb. 1973. Space, Time, Semantics, and the Child. In Timothy E. Moore (ed.) *Cognitive Development and the Acquisition of Language*. New York: Academic Press.
- Emmorey, Karen. In Press. *Language, Cognition and the Brain: Insights from Sign Language Research*. Lawrence Erlbaum.
- Fillmore, Charles. 1968. The Case for Case. In Emmon Bach & Robert T. Harms (eds.) *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston.
- Gruber, Jeffrey S. 1965. Studies in Lexical Relations. PhD dissertation, MIT. Reprinted as part of *Lexical Structures in Syntax and Semantics*, 1976. Amsterdam: North-Holland.
- Herskovits, Annette. 1982. Space and the Prepositions in English: Regularities and Irregularities in a Complex Domain. PhD dissertation, Stanford University.
- Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge, MA: MIT Press.
- Leech, Geoffrey. 1969. *Towards a Semantic Description of English*. New York: Longman Press, 1969.
- Liddell, Scott. Forthcoming. Sources of Meaning in ASL Classifier Predicates. In Karen Emmorey (ed.) *Perspectives on Classifier Constructions in Signed Languages* (provisional title). Likely publisher: Cambridge University Press
- Talmy, Leonard. 1983. How Language Structures Space. In Herbert L. Pick, Jr. & Linda P. Acredolo (eds.) *Spatial Orientation: Theory, Research, and Application*. New York: Plenum Press.
- Talmy, Leonard. 2000a. *Toward a Cognitive Semantics*, volume I: *Concept Structuring Systems*. Cambridge, MA: MIT Press.
- Talmy, Leonard. 2000b. *Toward a Cognitive Semantics*, volume II: *Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press.
- Talmy, Leonard. In press. The representation of Spatial Structure in Spoken and Signed Language. In K. Emmorey (Ed.), *Perspectives on Classifier Constructions in Sign Languages*. Mahwah, NJ: Lawrence Erlbaum and Associates.

Leonard Talmy

Zubin, David & Soteria Svorou. 1984. Orientation and Gestalt: Conceptual Organizing Principles in the Lexicalization of Space. With S. Choi. In David Testen, Veena Mishra & Joseph Drogo. *Lexical Semantics*. Chicago: Chicago Linguistic Society.

## **Pronoun Borrowing**

SARAH G. THOMASON and DANIEL L. EVERETT

*University of Michigan and University of Manchester*

### **0. Introduction**

A recurring theme in theoretical discussions of language contact is the question of borrowability—specifically, whether there are any substantive constraints governing the kinds of lexicon and structure that can be borrowed. Nowadays historical linguists are less likely to propose absolute constraints than they used to be, because everyone knows at least a few examples of ‘odd’ borrowings. Still, the feeling that some things ought not to be borrowed persists among both historical linguists and specialists in language contact.

Perhaps the most commonly mentioned hard-to-borrow lexical feature is the category of personal pronouns. The reasoning, usually implicit, seems to be roughly this: personal pronouns comprise a closed set of forms situated between lexicon and grammar; they form a tightly structured whole and are so deeply embedded within a linguistic system that borrowing a new personal pronoun, and in particular a new pronominal paradigm, would disrupt the workings of the system. Therefore, the argument goes, it is extremely unlikely that any language (or rather any speech community) would borrow pronouns.

This paper presents evidence that, given appropriate social circumstances, pronouns and even whole pronominal paradigms are readily borrowed. It is certainly more difficult to find examples of borrowed personal pronouns than examples of borrowed nouns denoting newly acquired cultural items; still, pronoun borrowing is nowhere near as rare as one would suppose from reading the literature. The larger goal of the paper is to provide one more piece of evidence in support of the claim that that speakers’ deliberate choices may be the most important factor motivating the borrowing of ‘hard-to-borrow’ features (see Thomason 1999). And this in turn leads to the conclusion that it is rash, in attempts to untangle a complex historical picture, to treat pronominal paradigms as safe markers of genetic linguistic inheritance.

### 1. Claims about Pronoun Non-borrowing: A Few Examples

A fairly typical, and typically cautious, textbook statement about borrowability is McMahon's formulation (1994:204):

'Some words also seem to be more borrowable than others: specifically, basic vocabulary...is only infrequently affected, and then almost always in situations where neither of the languages involved is perceived as more prestigious than the other. English borrowed a good deal of basic vocabulary, including *skin*, *sky*, *get* and the pronouns *they*, *them*, *their*, from Norse in the late Old and early Middle English period. . .'

McMahon doesn't explain the basis of her belief that an absence of social asymmetry is a major factor in facilitating the borrowing of basic vocabulary, but it may be because her main example, English borrowings from Norse, is presumed to have been such a situation. Prestige is in any case a slippery concept; we believe that the relevant social factors are far more complex. (Similarly, the general reasons for lexical borrowing are by no means confined to the usual suspects, 'need' and 'prestige'.)

Stronger statements about borrowability have been made by other scholars, perhaps most notably, in recent years, by R.M.W. Dixon and Johanna Nichols & David Peterson. Dixon, while acknowledging that individual pronouns are sometimes borrowed, says that 'there are certain grammatical phenomena that are very unlikely to be borrowed, under any circumstances. These are:...complete paradigms, e.g. a pronoun paradigm, a noun declension or a verb conjugation' (1997:22). Nichols & Peterson, though their main focus is also on paradigms, argue more generally that 'barring non-normal transmission, pronouns are almost always inherited...the cases where pronouns are known to have been transferred from one language to another are generally not routine borrowing' (1996:337-38; they do not define 'routine borrowing'). They reiterate this argument in their response to criticisms by Campbell (1997), claiming that '...pronouns are almost always inherited and almost never borrowed' (1998:610).

Dixon and Nichols & Peterson have theoretical reasons for believing that pronoun borrowing is vanishingly rare. For Dixon, the significance is that, since pronominal paradigms are (in his view) among the linguistic features that are least likely to be borrowed, they are one of 'the surest indicators of genetic relationship' (1997:22). Nichols & Peterson are less clear in their theoretical argumentation, but they too seem to be suggesting that, because pronouns are so rarely borrowed, a pattern of paradigmatic pronominal similarities over several different language families probably indicates distant genetic relatedness (1996:337-38).

Linguists interested in developing claims of long-range genetic relationships very often agree with Dixon in considering pronouns to be safe indicators of genetic relatedness. Vitaly Shevoroshkin, for instance, says that 'pronouns of the sort "I", "me", "thou", "thee" are not borrowed from language to language; they are inherited' (1989:6), and Joseph Greenberg & Merritt Ruhlen, focusing on pronominal

affixes rather than independent pronouns, claim that ‘pronominal affixes are among the most stable elements in languages: they are almost never borrowed’ (1992:97—cited, along with the Shevoroshkin quotation, in Campbell 1997:340-341).

## **2. Some Examples of Borrowed Pronouns**

A search of the literature, especially for Southeast Asia and the Pacific but also in the Americas and elsewhere, turns up a sizable number of examples of borrowed pronouns. Several are cited in Campbell’s response to Nichols & Peterson 1996 (1997:340): besides English *they*, *their*, *them*, there are the independent pronouns of Miskito (Nicaragua), borrowed from Northern Sumu (Campbell cites Kenneth Hale, p.c., as the source of this information); and Alsea (Oregon), a non-Salishan language, has apparently borrowed ‘a whole set of Salishan pronominal suffixes’ (citing Kinkade 1978).

Campbell also refers to documented cases of borrowed pronouns in ‘Southeast Asian languages, Austronesian, and Papuan languages’ (1997:340). As an example he cites, from Foley (1986:210), the striking case of two apparently unrelated non-Austronesian (so-called ‘Papuan’) languages of New Guinea that share first- and second-person pronouns, but with reversed meanings: Kambot borrowed the Iatmul word for ‘I’ in the meaning ‘you’, and the Iatmul word for ‘you (feminine)’ as ‘I’. The phenomenon of pronoun borrowing with reversed meanings may not be as rare as one might expect. Miller, writing about pronouns in Japanese and Altaic languages, writes that

‘one of the most perplexing problems...[is] the shifting back and forth in semantic category from one person to another...Common to all these examples of semantic interchange between first-person and second-person are two sociolinguistic elements—the self-deprecatory employment of a pejorative second-person in the resultant sense of a humble first-person; and the converse employment of a humble, self-deprecatory first-person in the sense of a particularly pejorative second-person’ (1971:173).

Now, Miller’s goal is to argue forcefully for a genetic relationship linking the Altaic languages (a genetic unit that is itself controversial) to Japanese and Korean. But his picture of the semantic shifts in first- and second-person pronouns resembles the pattern in the shared pronouns of Kambot and Iatmul so closely that it seems more likely to support an alternate hypothesis, namely, that lexicon and structure are shared by the proposed Altaic languages because of long and intimate contact rather than because of inheritance from a common ancestor. We have not carried out any kind of systematic analysis of the Japanese, Korean, Tungusic, Mongolian, or Turkic data; the point of this observation is therefore not to argue for an areal source for the semantic interchange in Altaic pronoun systems, but rather to argue against the too-easy assumption that the partly shared Altaic pronominal systems must be inherited.

Borrowed individual pronouns are fairly easy to find in lists of loanwords from a variety of languages. To take just one example, Thurgood’s list of loanwords

into Proto-Chamic (Austronesian; Southeast Asia) includes two personal pronouns, *\*dahla?* ‘I (polite)’ (from an unknown source) and *\*ha* ‘you; thou’ (from an unknown source); and his list of loanwords into Chamic languages after the breakup of Proto-Chamic includes two more pronouns, *biŋ* ‘we’ (from Mon-Khmer) and *ih* ‘you; thou’ (from Mon-Khmer) (1999:337, 338, 351, 356).

In some cases, pronouns seem to have been borrowed to fill a perceived gap in the pronominal paradigm. One example is found among North Halmaheran languages (non-Austronesian; north Moluccas, Indonesia), which are in close contact with Austronesian languages that have an inclusive/exclusive ‘we’ distinction. Voorhoeve (1994:661) says that the first person plural exclusive pronoun in all these languages is probably of Austronesian origin, although the first person plural inclusive pronoun seems to be a native form. Conversely, some varieties of Malay in Indonesia have lost their inherited exclusive/inclusive ‘we’ distinction through borrowing from non-Austronesian languages—that is, they have borrowed a structural pronominal pattern. An example is Manado Malay (in Sulawesi), which has not only lost this distinction but has borrowed both its singular and its plural second-person pronouns from Ternate (Prentice 1994:423); this variety of Malay has thus restructured its pronominal paradigm to match that of the non-Austronesian language Ternate more closely. The same thing has apparently happened in other varieties of Malay as well, under the influence of other languages that lack inclusive vs. exclusive ‘we’—the distinction has been lost, and first- and/or second-person pronouns have been borrowed, from such languages as Hokkien (Min) Chinese, Portuguese, and English (Smith & Donohue 1998). Smith & Donohue comment on the fact that ‘varieties of Malay are prone to borrow pronouns from other sources’.

Yet another example comes from the Philippines, where the Spanish-based creole Chavacano is spoken. In Zamboanga city, according to Forman (2000), young people often speak Chavacano in a way that their elders frown upon; this ‘way of talking is characterized by the preference or tendency to insert Tagalog pronouns, in particular the second person singular, into otherwise Zamboangueno utterances’—a pattern that apparently arose because of the influx of large numbers of young soldiers into the city. It’s not clear whether this pattern reflects fixed borrowing of the pronouns; but it clearly isn’t as ephemeral as one would expect with ordinary code-switching. The recently published *Chabacano de Zamboanga handbook and Chabacano-English-Spanish dictionary* (Camins 1999) lists alternate sets of plural pronouns, one of Spanish origin and one of Tagalog origin: 1pl inclusive *Nosotros*; *Kita*, 1pl exclusive *Nosotros*; *Kame*, 2pl *Ustedes*, *Vosotros*; *Kamo*, 3pl *Ellos*; *Sila* (pp. 10, 84-85). Note that the Tagalog loanwords introduce an exclusive/inclusive ‘we’ distinction into Chavacano. According to Camins, the usage of these different pronoun sets depends on status relations between speaker and hearer. He identifies three status categories, formal (to express respect for age or higher status), familiar (to express familiarity or same status as speaker), and common (to express same status or lower status) (p. 10). The Spanish-origin set

### *Pronoun Borrowing*

is used for formal address and, in 2pl only, for familiar address; the Tagalog-origin set is used for common address and, in 1pl at least, for familiar address. (It's not entirely clear from Camins' account whether familiar 3pl address requires Spanish- or Tagalog-origin pronouns.)

Finally, sets of pronominal affixes are sometimes borrowed. In Meglenite Rumanian, for instance, the fully inflected verb forms 1sg *aflu* 'I find' and 2sg *aflī* 'you find' have been augmented by the addition of Bulgarian person/number/tense suffixes, 1sg present *-m* and 2sg present *-š*, to yield double-marked forms *aflum* and *aflīš* (Sandfeld 1938:59). Dawkins, analyzing the heavy Turkish influence on dialects of Asia Minor Greek almost a century ago, reports that Turkish 1pl and 2pl suffixes have been borrowed into some dialects of Silli and Cappadocia and added to Greek verbs (1916:59, 144; these borrowed Turkish pronominal suffixes may be optional—Hovdhaugen 1976:148). Both the Rumanian and the Greek examples involve partial paradigms rather than entire paradigms, and it's futile to speculate about why speakers of one language borrowed only singular affixes while speakers of another borrowed only plural affixes. In any case, they at least attest to the possibility of borrowing sets of pronominal affixes.

An especially striking example is Mednyj Aleut, a bilingual mixed language with both Aleut and Russian grammatical subsystems. Mednyj Aleut apparently developed in the 19th century on Mednyj (Copper) Island off the northeastern coast of Russia, probably among the mixed-blood offspring of Russian fathers and Aleut mothers. The base language is clearly Aleut, although there are many Russian loanwords and some syntactic influence from Russian as well. But according to the first modern linguistic analysis of Mednyj Aleut, (Menovščikov 1969), the entire finite verbal inflection of Aleut—an enormously complex system—has been replaced by Russian verb inflection. The borrowed inflection includes person/number/tense suffixes and also, because Russian past-tense forms were originally participles and therefore do not code person, independent Russian pronouns in the past tense only. Non-finite verb inflection is still Aleut. (For further discussion, see Thomason & Kaufman 1988:233-238, Golovko & Vakhtin 1990, Golovko 1994, Thomason 1997).

The relevance of the Mednyj Aleut case to a discussion of pronoun borrowing has been disputed. Nichols & Peterson dismiss it on the ground that Mednyj Aleut is a mixed language (true) and that '[i]t is not that Russian personal endings are borrowed into Copper Island Aleut, but rather they are native to one of the two ancestral languages' (1996:338); similarly, in their response to Campbell, they claim that Russian pronominals in Copper Island Aleut 'are inherited there', because 'Russian is one of the two ancestors' (1998:610). The conception of genetic relationship and inheritance from a common ancestor that these quotations reveal does not match any standard historical linguistic view, and Nichols & Peterson do not explain just what they have in mind by using the term 'inherited'. Like some other bilingual mixed languages, Mednyj Aleut was created in a rather short period

of time by bilinguals—the creators must have been bilingual to a considerable extent, because the amount of distortion in either component of the mixed language is small. The base language is clearly Aleut: the bulk of the lexicon, together with the entire nominal inflectional system and all the non-finite verbal inflection, is Aleut. So there is no question of inheritance from Russian; instead, the Russian forms were imported into an Aleut matrix. In any case, for excellent methodological reasons, standard notions of genetic relationship do not permit positing more than one ancestor: either a language's major subsystems are all descended primarily from a single ancestor or they aren't; if they aren't, as with Mednyj Aleut and other mixed languages, then the language has no ancestors at all in the historical linguist's technical sense (see Thomason & Kaufman 1988 for discussion). This is not to say, of course, that there are no source languages that contributed to its lexicon and structure. Nichols & Peterson are correct in attributing the Russian inflection in Mednyj Aleut to 'non-normal transmission' (1996:338), but of course that's true for every specific example of borrowed lexicon and structure in every language, including languages that fit comfortably into a traditional family tree: borrowed material is by definition not inherited from the parent language. It's also true that Mednyj Aleut was created, probably deliberately, to serve as a symbol of the new socioeconomically distinct group of mixed-bloods. But if those bilinguals could combine an inflectional system, including pronominals, with another language's grammar, other bilinguals could do the same thing. And, as we've shown in this section, less dramatic but still striking instances of pronominal borrowing can be adduced from a variety of other languages too.

Moreover, new evidence suggests the sort of process by which the startling mixture found in Mednyj Aleut might have arisen. In an article containing some very interesting examples of interference from Russian in Tungusic languages, Malchukov (2003) notes one set of features adopted by Evenki, a North Tungusic language, from the Turkic language Yakut: Yakut interference in Evenki includes the paradigm of the volitional mood, which comprises a mood suffix borrowed from Yakut together with an entire set of borrowed Yakut personal endings. Why this borrowing should occur in a non-'basic' mood category is unclear; however, it suggests a possible route by which Russian inflectional affixes could have made their way into the Mednyj Aleut finite verb, at first replacing Aleut suffixes only in a secondary paradigm and only later spreading throughout the system of finite verb inflection. Unfortunately, if there were such intermediate stages in the development of Mednyj Aleut, we have no documentation of them. Malchukov's Evenki example does show, however, that the borrowing of finite verb inflection, including pronominal affixes, is not unique to Mednyj Aleut.

It probably isn't accidental that most of the examples mentioned above come from the Pacific and nearby Southeast Asia. In this part of the world pronouns do indeed appear to be borrowed quite readily; Wallace observes that in 'some parts of Southeast Asia,...personal pronouns and elements which have become personal

### *Pronoun Borrowing*

pronouns have moved from one language to another in relatively great numbers and with relatively great ease, sometimes replacing an indigenous set, sometimes expanding it' (1983:575). And Foley has said that pronouns in non-Austronesian languages of New Guinea are 'definitely not immune to borrowing, nor even particularly resistant' (1986:211); his Kambot/Iatmul example is just one example of the borrowing of partial pronominal paradigms. He comments that in Southeast Asia 'pronouns have already been shown to be prone to borrowing', citing among other examples the borrowing of English *I* and *you* in Thai and Indonesian (1986:210). Thai also has other borrowed pronouns: Christopher Court says that Thai boys often use the Chinese pronouns for 'I' and 'you' when speaking to each other (1998), and Samang Hiranburana reports that the complex set of Thai 'royal' pronouns has been claimed as loanwords from Khmer dating from the 14th and 15th centuries (1998).

Here a very interesting distinction is suggested by Court between 'closed' pronoun systems like those in European languages, where the general pattern is just one pronoun for a given person/number combination, and 'open' pronoun systems like those in Southeast Asian languages, where there may be (for instance) dozens of ways to say 'I' and 'you' (1998). Unsurprisingly, pronoun borrowing is more common in languages with 'open' systems (which may, from one perspective, be compared to the several sets of pronominal morphemes in a polysynthetic language like Montana Salish or Aleut). Similarly, and also with reference to Southeast Asian and nearby Pacific languages, Wallace argues that basing theories of pronominal stability on Indo-European languages, where pronominal systems have indeed been relatively stable, is 'founded on limited views of human social organization and cultural contact' and is therefore 'misguided' (1983:575).

For this paper, the crucial point in all these cases is that social factors, not linguistic ones, determine the likelihood of pronominal borrowing. If speakers want to borrow one pronoun or a whole set of pronouns, they can do so; and sometimes speakers do want to do this. The borrowed pronouns may change the structure of the pronominal system significantly, as when a new category of inclusive vs. exclusive 'we' is introduced or lost through borrowing. Or they may merely alter the nature of social discourse, as when Thai speakers use English pronouns, which enable them to sidestep the traditional pronominal coding of social differences in such features as age, status, and degree of intimacy. Borrowing of this type cannot be predicted from general principles, any more than it can be ruled out on the basis of any general principles; it occurs if and only if speakers decide to make it happen—just as other lexical borrowing depends on speakers' decisions. It's worth noting in this context that some languages, including many of those in the Pacific Northwest Sprachbund of the U.S. and western Canada, have only a tiny handful of loanwords from English in spite of a hundred and fifty years of close contact with English. In other words, extensive lexical and structural borrowing is neither inevitable nor impossible in the most intense contact situations.

### 3. A Case Study: The Pronouns of Pirahã

Skepticism about the probability, or even the possibility, of pronoun borrowing has led some scholars to reject analyses of borrowed pronominal paradigms out of hand. In our opinion, this view leads to unwarranted conclusions about historical developments in specific languages and language families: if sets of pronouns are sometimes borrowed, then a language's pronouns cannot be automatically assumed to be 'fossils', relics that point directly to a language's genetic affiliation.

In this section we will discuss one case in some detail, to illustrate the kind of argumentation that's needed to make a convincing case for the borrowing of pronominal paradigms when external historical information is largely lacking. The case is that of Pirahã, an Amazonian language spoken in Brazil. According to the *Ethnologue*, which gives the language name as Múra-Pirahã, the language has about 150 speakers, who are 'quite monolingual' (Grimes 1992:31). The *Ethnologue* notes that it is probably related to the extinct language Matanawi; no wider genetic affiliations for the language have been established. Pirahã is the only surviving dialect of the language. As far as Everett has been able to determine, during visits to Mura settlements and discussions with anthropologists and government employees, the Mura dialect has not been spoken among the Muras for over two generations; the Muras have shifted to Portuguese.

The proposal that the entire set of pronouns of Pirahã was borrowed came originally from Aryon Rodrigues (personal communication to Everett, 1978). Citing Rodrigues, Everett discussed the idea briefly in three articles (1979, 1986, in press). In this section we offer more detailed arguments to support the proposal that Pirahã pronouns were borrowed from Tupí-Guaraní, either from Tenharim or from Nheengatu, and either directly or indirectly, via Mura.

In 1978, at the beginning of more than twenty-five years of fieldwork on Pirahã (including nearly six full years in residence in the village), Everett discussed some of his data with his M.A. thesis advisor, Rodrigues. Upon seeing the Pirahã pronouns, Rodrigues noted that they are nearly identical to the singular Nheengatu forms—in fact, that they are nearly identical to the Proto-Tupí-Guaraní forms. Rodrigues was quite correct, as we will show. We will argue that Pirahã pronouns were most likely borrowed from one (or both) of the Tupí-Guaraní languages with which Pirahã speakers are known to have been in contact, Nheengatu and Tenharim, but that the pronouns probably entered Pirahã indirectly, via Mura, rather than directly from one of these languages.

Pirahã has only three basic personal pronouns (Table 1). There are also several pronominal clitics that are shortened forms of longer (non-pronominal) words; we list the two most common of these in Table 1, but we will not discuss them further.

## *Pronoun Borrowing*

TABLE 1. **Pirahã pronouns.**

| Phonemic shape | Phonetic shape | gloss                  |
|----------------|----------------|------------------------|
| /ti/           | [č̥i]          | ‘1st person’           |
| /gi/, /gia/    | [ni], [niʔa]   | ‘2nd person’           |
| /hi/           | [hi]           | ‘3rd person’           |
| /ʔi/           |                | ‘3rd person feminine’  |
| /ʔis/          |                | ‘3rd person non-human’ |

The last two of these five pronominal forms, /ʔi/ ‘3rd fem’ and /ʔis/ ‘3rd non-human’ have a special status: neither form is used in isolation. So, for example, in response to the question, ‘Who did that?’, one could answer /ti/ ‘me’, /gi/ ‘you’, or /hi/ ‘him’, but one could not answer /ʔi/ ‘she’ or /ʔis/ ‘the animal’. To refer to a woman or animal actor, it is necessary to use the full form of the word from which the clitic is derived, /ʔipoihi/ ‘woman’ or /ʔisi/ ‘animal’. (The third-person pronoun /hi/ has specifically masculine reference only when it is contrasted with the 3rd feminine clitic /ʔi/; see D. Everett 1986.)

The three basic Pirahã pronouns comprise one of the simplest pronominal systems known. They are often optional in discourse, so that their functional load is not as great as that of pronouns in many other languages—especially given the fact that Pirahã has no form of agreement marked on the verb, aside from the pronominal clitics. Note that there is no singular/plural distinction in Pirahã pronouns; in fact, the language has no number distinctions of any kind in its grammar. The pronouns are all number-neutral. If speakers want to talk about more than one of something, they use the quantifier word /ʔogiaágaó/ ‘all’, which can combine with, for instance, /ti/ to mean ‘we (all) go’. Note also that the pronouns form a syntactic class of words separate from nouns; they act like clitics and can double nouns (see D. Everett 1987).

In considering the proposal that the Pirahã pronouns are loanwords, only the three basic pronouns /ti/, /gi/, and /hi/ are relevant, since they are the only ‘pure’ pronouns and the only pronominal forms that can function as independent pronouns. And in comparing Pirahã pronouns to Tupí-Guaraní pronominals, it’s vital to take the entire inventory of Tupí-Guaraní pronouns into account, because the languages of this family have two sets of pronouns each. For Nheengatu, for instance, some sources, e.g. Tastevin (1910:62), give only one of the language’s two sets, and it’s not the set that matches the Pirahã pronouns; other sources, e.g. Gonçalves Dias (1965:29, 47, 69), give both sets. The relevant set is the independent ergative pronominal paradigm of Nheengatu (and, according to Jensen 1998, this was also the Proto-Tupí-Guaraní paradigm); these are the most frequently-occurring pronouns in the language:

TABLE 2. **Nheengatu free ergative pronouns.**

| Phonemic shape | Phonetic shape | gloss           |
|----------------|----------------|-----------------|
| /xe/           | [šɪ]           | ‘1sg’           |
| /ne/           | [ne], [nde]    | ‘2sg’           |
| /ahe/          |                | ‘3sg/pl’        |
| /iandé/        |                | ‘1pl inclusive’ |
| /oré/          |                | ‘1pl exclusive’ |
| /pe/, /pee/    |                | ‘2pl’           |

Another relevant form is the prefix (or clitic) /i-/ [ɪ], [e] ‘3sg/pl’.

Compare this Nheengatu set to the very similar pronouns of Tenharim (from Helen Pease, p.c. 1998):

TABLE 3. **Tenharim free ergative pronouns.**

| Phonetic shape | gloss                          |
|----------------|--------------------------------|
| [jɨ]           | ‘1sg’                          |
| [nde], [ne]    | ‘2sg’                          |
| [hea]          | ‘3sg feminine’                 |
| [ahe]          | ‘people’, or ‘person now dead’ |

Our claim is that the basic Pirahã pronouns are nearly identical to those of Nheengatu and Tenharim. Superficially, however, the Pirahã pronouns don’t look much like the Tupí-Guaraní pronouns; so this proposal will not be convincing without some additional information about the phonology of Pirahã that shows how the phonetic realizations of the Tupí-Guaraní forms align with the Pirahã phonemic system.

Pirahã has just eight consonants in the segmental inventory of men’s speech, and seven in women’s speech: /p, b, t, k, g, ʔ, h/ and, in men’s speech only, /s/ (see K. Everett 1998 for a phonetic study of Pirahã segments and prosodies). Women substitute /h/ for men’s /s/. Several consonant phonemes have significant allophonic variation; for our purposes, the relevant allophones are [b] and [m] for /b/; [g] and [n] for /g/; [t] and [č] for /t/; and [s] and [š] for /s/. The two alveopalatal allophones, the affricate [č] and the fricative [š], occur always and only before a front vowel; but [š] of course occurs only in men’s speech. Both men and women have three vowel phonemes, front, central, and back: /i/, with allophones [ɪ] and [e]; /a/; and /u/, with allophones [u] and [o].

Now, compare the Pirahã pronouns to the Nheengatu pronouns. Nheengatu 1sg /xe/ is pronounced [šɪ], according to various sources (and Aryon Rodrigues, p.c. to Everett, 1998). The only alveopalatal phones in Pirahã are [č] and [š]; but since [š] is not found in women’s speech, [č] is the only alveopalatal consonant found in both men’s and women’s speech. This makes [č] the most likely Pirahã nativization of Nheengatu [š]. The affricate [č] would be even more likely if the source pronoun

### *Pronoun Borrowing*

were instead Tenharim [jɪ] ‘1sg’; note also that, according to Jensen (1998:6), the relevant Proto-Tupí-Guaraní pronoun began with a voiceless alveopalatal affricate.

The Nheengatu 2sg pronoun varies between [nde] and [ne]; the second pronunciation is conditioned by a preceding nasal segment. Since Pirahã has no [d], but does have [n] as the word-initial allophone of /g/, both [nde] and [ne] would be expected to be borrowed as Pirahã /gi/ [ni]. (See D. Everett 1979 for details of /g/ allophony.)

This leaves the Pirahã third-person pronoun /hi/ to be accounted for. Both Nheengatu and Tenharim have a third-person pronoun /ahe/, though with slightly different meanings (‘3sg/pl’ in Nheengatu, ‘people’ or ‘person now dead’ in Tenharim). In addition, Nheengatu has a prefix (or clitic) form /i-/ [ɪ, e] also meaning ‘3sg/pl’. Now, Pirahã lacks vowel-initial syllables entirely (D. Everett 1988, K. Everett 1998), so that the Nheengatu form /i-/, if borrowed into Pirahã, would need an added prothetic consonant—presumably either the unmarked continuant /h/ or the unmarked stop /ʔ/—to satisfy the language’s syllable structure constraints (D. Everett 1988). In this instance, a prothetic /h/ seems the more likely choice, because a prothetic glottal stop would make the general third-person pronoun homophonous with the derived Pirahã clitic /ʔi/ ‘3 feminine’ (if this feminine clitic already existed in Pirahã at the time of borrowing). Pirahã /hi/ is also a reasonable nativization of the Nheengatu (or Tenharim) third-person pronoun /ahe/; since this pronoun already has a consonant, deleting the initial vowel rather than adding a second consonant would not be surprising. The Pirahã pronoun /hi/ and the Nheengatu pronoun /ahe/ share a striking, and unusual, usage feature which adds strength to the case for a historical connection between them: in addition to their use as ordinary third-person pronouns, both are also used as demonstratives—even for non-third persons, as in Pirahã *hi ʔobaaʔai ti* ‘I am really smart’, literally ‘This one/someone sees well, me’).

We do not claim to have demonstrated in this section that the three basic pronouns of Pirahã were borrowed from Nheengatu or Tenharim, or (conceivably) both, either directly or indirectly by way of Mura (see below). What we have demonstrated is that Pirahã pronouns match the relevant pronoun sets of Nheengatu and Tenharim very closely, phonologically and, in the case of /hi/ in one quite specific usage feature. The match is so close, in fact, that coincidence is not an appealing explanation, though with such short forms it is still a possibility. As we noted above, borrowing is in itself quite likely, because the Pirahãs have had close long-term contacts with speakers of both Nheengatu and Tenharim—especially with Nheengatu, which was for centuries the trade language of Amazonia. And, as we saw in earlier sections, pronominal borrowing is not especially unusual under certain kinds of social circumstances.

In spite of the close contacts between Pirahã, Nheengatu, and Tenharim, we consider it most probable that Pirahã acquired its personal pronouns from its sister dialect Mura rather than directly from Nheengatu and/or Tenharim. The Pirahãs

have long resisted influence from outside cultures and their languages; their language has relatively few of the more usual sorts of loanwords, though there are several from Nheengatu, among them *purasey* ‘to dance’, *cosisi* ‘to sleep’ – which Nheengatu probably originally borrowed from Portuguese *coxilar* ‘to nap’ – and *maoa* ‘dead’. Pirahã also has a few words borrowed directly from Portuguese, e.g. *boitohoi* ‘motor’ (a slang word for boat in that region) and *nahiao* ‘airplane’, and one or two from English (e.g. *topagaha* ‘tape-recorder’). But the Pirahãs and Muras used to be in daily contact, and it is plausible that their very close contacts, and the mutual intelligibility between their languages, motivated and facilitated the transfer of pronouns from Mura to Pirahã. There is solid evidence that Mura personal pronouns were themselves borrowed from Nheengatu (also called the *Lingua Gerãl*, or ‘general language’), as reported by Nimuendaju, who reported on Mura toward the end of its existence (1948:257):

‘Martius’ contention that most of the words of the Mura language are of Tupian origin has remained unsubstantiated. Even the number of elements adopted from the *Lingua Geral* is strangely small. Most noticeable are the regular use of the first and second singular, personal pronouns, and first person plural of *Lingua Geral*.’

In other words, even as the Muras were shifting rapidly to Portuguese, they apparently shared with the Pirahãs the social characteristic of resisting borrowing; and among the few words they borrowed were at least two of the very pronouns that, we propose, Pirahã likely borrowed, and ultimately from the same source. We have only a tantalizingly incomplete picture of contacts between Muras and Nheengatu speakers, between Pirahãs and Nheengatu (and Tenharim) speakers, and even between Muras and Pirahãs. What we do know about these contacts makes our hypothesis plausible; and the close match in structure and function of the Pirahã and Nheengatu pronouns seems to support it. (Unfortunately, no such detailed analysis is possible for Mura, because this dialect was not extensively documented before it died.)

Caution is required, of course: as noted, we have too little information about the specific social circumstances of the relevant contacts, and we have much too little information about the history of Pirahã, given its lack of well-attested relatives. The language does, or did, have relatives, including at least Matanawi, Yahahi, and Bohura, as well as Mura, which was a dialect of the same language as Pirahã; but all these relatives are extinct, and we have virtually no linguistic data for them. That is, we can establish two of the requisites for a successful argument for borrowing in this case: there was certainly extensive contact, and the pronouns in question are certainly old in Tupí-Guaraní languages. It is also true that no genetic relationship has been established between Pirahã and Tupí-Guaraní. Still, we can’t prove that the pronouns in question are innovative in Pirahã, or, for that matter, in Mura, in spite of Nimuendaju’s observation; and there are few definite loanwords in Pirahã from Tupí-Guaraní or any other language. Nevertheless, even with large gaps in the

case for borrowing, on balance it seems to be the best historical explanation for the Pirahã facts.

#### **4. Conclusion**

The rich body of evidence from Southeast Asia, New Guinea, and parts of the Pacific, together with scattered examples from other parts of the world, supports our claim that pronoun borrowing—like other contact phenomena—is subject to deliberate and conscious choices made by speakers of a variety of languages. These choices in turn are conditioned by a variety of cultural traits. But since cultural traits such as attitudes toward borrowing are extremely difficult to discover in past contact situations, it is never safe to assume any particular cultural stance toward borrowing, without external evidence. And this makes the common assumption that pronouns must be inherited because they are almost never borrowed unwarranted at best: we can't know, without specific linguistic or social evidence, whether a given past culture was more like those of Indo-European languages, where most pronouns are in fact inherited, or more like those of languages whose speakers borrow pronouns freely. In other words, assuming as a default that matching sets of pronouns infallibly indicate genetic relationship is unwise: there are no shortcuts to the establishment of genetic relationship.

#### **References**

- Camins, Bernardino S. 1999. *Chabacano de Zamboanga handbook and Chabacano-English-Spanish dictionary*. Zamboanga City: Office of the City Mayor.
- Campbell, Lyle. 1997. Amerindian Personal Pronouns: A Second Opinion. *Language* 73:339-51.
- Court, Christopher. 1998. Untitled posting on the SEALTEACH list, 30 April 1998.
- Dawkins, R.M. 1916. *Modern Greek in Asia Minor*. Cambridge: Cambridge University Press.
- Everett, Daniel L. 1979. Aspectos da fonologia do Pirahã. M.A. thesis, Universidade Estadual de Campinas, Campinas, Brazil.
- Everett, Daniel L. 1986. Pirahã. In Desmond Derbyshire and Geoffrey Pullum, eds., *Handbook of Amazonian Languages I*, 200-326. Berlin: Mouton de Gruyter.
- Everett, Daniel L. 1987. Pirahã Clitic Doubling. *Natural Language and Linguistic Theory* 5:245-76.
- Everett, Daniel L. 1988. On metrical constituent structure in Pirahã phonology. *Natural Language and Linguistic Theory* 6:207-46.
- Everett, Daniel L. In press. A Critical Evaluation of Greenberg's Classification of Arawan, Chapacuran, and Muran. In Allan Taylor, ed., *Language and Prehistory in the Americas*. Stanford: Stanford University Press.

- Everett, Keren M. 1998. The Acoustic Correlates of Stress in Pirahã. *Journal of Amazonian Linguistics* 1/2:104-62.
- Foley, William A. 1986. *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
- Forman, Michael L. 2000. Confidence in Chabacano: Counterbalance to a Western Ideology of Language. Paper presented at the conference 'Shedding Light on the Chabacano Language', Ateneo de Manila University, 19-20 October 2000.
- Golovko, Evgenij V. 1994. Copper Island Aleut. In Peter Bakker and Maarten Mous, eds., *Mixed Languages*, 113-21. Amsterdam: Uitgave IFOTT.
- Golovko, Evgenij V., and Nikolai B. Vakhtin. 1990. Aleut in Contact: The CIA Enigma. *Acta Linguistica Hafniensia* 72:97-125.
- Gonçalves Dias, Antonio. 1965. *Dicionário da língua tupí: Chamada Língua Geral dos índigenas do Brasil*. Rio de Janeiro: Livraria São José.
- Greenberg, Joseph H., and Merritt Ruhlen. 1992. Linguistic Origins of Native Americans. *Scientific American* 267.5:94-99.
- Grimes, Barbara F., ed. 1992. *Ethnologue: Languages of the World*. 12th edn. Dallas: Summer Institute of Linguistics.
- Hiranburana, Samang. 1998. Untitled posting on the SEALTEACH list, 22 April 1998.
- Hovdhaugen, Even. 1976. Some Aspects of Language Contact in Anatolia. *University of Oslo Working Papers in Linguistics* 7:142-60.
- Jensen, Cheryl. 1998. The Use of Coreferential and Reflexive Markers in Tupí-Guaraní Languages. *Journal of Amazonian Linguistics* 1/2:1-49.
- Kinkade, M. Dale. 1978. Alsea Pronouns. Paper presented at the 17th Conference on American Indian Languages, American Association of Anthropology, Los Angeles.
- Malchukov, Andrei L. 2003. Russian Interference in Tungusic Languages in an Areal-Typological Perspective. In P. S. Ureland, ed., *Studies in Eurolinguistics*, vol. 1: *Convergence and Divergence of European Languages*, 235-251. Berlin: Logos Verlag.
- McMahon, April M.S. 1994. *Understanding Language Change*. Cambridge: Cambridge University Press.
- Menovščikov, G.A. 1969. O nekotoryx social'nyx aspektax èvoljucii jazyka. In *Voprosy social'noj lingvistiki*, 110-34). Leningrad: Nauka.
- Nichols, Johanna, and David A. Peterson. 1996. The Amerind Personal Pronouns. *Language* 72:336-71.
- Nichols, Johanna, and David A. Peterson. 1998. Amerind Personal Pronouns: A Reply to Campbell. *Language* 74:605-14.
- Nimuendaju, Curt. 1948. The Mura and Pirahã. In J. Steward, ed., *Handbook of South American Indians*, 255-269. Washington, DC: Bureau of Ethnology.
- Sandfeld, Kristian. 1930. *Linguistique balkanique: problèmes et résultats*. Paris: Librairie C. Klincksieck.

*Pronoun Borrowing*

- Shevoroshkin, Vitaly. 1989. A Symposium on the Deep Reconstruction of Languages and Cultures. In Vitaly Shevoroshkin, ed., *Reconstructing Languages and Cultures: Materials from the First International Interdisciplinary Symposium on Language and Prehistory*, 6-8. Bochum: Brockmeyer.
- Smith, John Charles, and Mark Donohue. 1998. What's Happened to Us? Developments in Malay Pronouns. *Oceanic Linguistics* 37.1:65-84.
- Tastevin, Constant. 1910. *La langue Tapihiya dite tupí ou Neêngatu (belle langue), grammaire, dictionnaire, et textes*. Vienna: Kaiserliche Akademie der Wissenschaften, Schriften der Sprachenkommission, vol. 2.
- Thomason, Sarah G. 1997. Mednyj Aleut. In Sarah G. Thomason, ed., *Contact Languages: A Wider Perspective*, 449-68. Amsterdam: John Benjamins.
- Thomason, Sarah. 1999. Speakers' Choices in Language Change. *Studies in the Linguistic Sciences* 29/2:19-43.
- Thomason, Sarah Grey, and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change*. (Oceanic Linguistics Special Publication no. 28.) Honolulu: University of Hawai'i Press.
- Voorhoeve, C.L. 1994. Contact-induced Change in the Non-Austronesian Languages in the North Moluccas, Indonesia. In Tom Dutton and Darrell T. Tryon, eds., *Language Contact and Change in the Austronesian World*, 649-74. Berlin: Mouton de Gruyter.
- Wallace, Stephen. 1983. Pronouns in Contact. In Frederic B. Agard, Gerald Kelley, Adam Makkai, and Valerie Becker Makkai, eds., *Essays in Honor of Charles F. Hockett*, 573-89. Leiden: E.J. Brill.

Sarah G. Thomason  
4080 Frieze Building  
University of Michigan  
Ann Arbor, MI 48109-1285  
thomason@umich.edu

Daniel Everett  
Department of Linguistics  
University of Manchester  
Manchester M13 9PL  
U.K.  
dan.everett@manchester.ac.uk



## Courtroom Accommodation: Lawyers' Use of Foreigner Talk

ELIZABETH CLOSS TRAU GOTT  
*Stanford University*

### 0. Introduction

There has been a great deal of research into courtroom language, especially lawyers' questioning of defendants and witnesses (see papers in Gibbons 1984, Danet 1985, Shuy 1993, Conley and O'Barr 1998, Tiersma 1999).<sup>1</sup> Some of this research has focused on communication failure due to cultural and linguistic difference (Gumperz 1982). Much has focused in addition on power disparities. Some of these disparities have been attributed to factors such as the social pressure of the courtroom situation, and to the use of complex legal language by legal professionals (Tiersma 1999). Others have been attributed to age, for example in the cross-examination of children (Brennan 1994, Brennan and Brennan 1988), or systematic differences in interactional styles and communication patterns, for example differences between men's and women's styles (Conley and O'Barr 1998), or between English and aboriginal language users in Australia (Eades 1994). In some instances they have also been attributed to lawyers' racism (Eades 1994). In this work the focus has been on the dysfunctionality and powerlessness of certain types of discourses from the perspective of the defendants, or on how judges maintain control over proceedings (Philips 1998).

A complementary, and apparently much more limited, line of research has explored the role of accommodation in legal discourse, with attention to convergences between the parties to increase intelligibility and efficiency in court procedures. For example, Linell (1991) investigated the extent to which lawyers in Swedish larceny and fraud trials modified their questions to increase length of responses by using intonation questions vs. inverted questions. Linell assumes that intonation questions, having declarative syntax, are less controlling; by

---

<sup>1</sup> I am grateful to Sarah Roberts for drawing my attention to the pre-trial transcript of the *People of California vs. Ah Jake* trial, which Randall Kim made available at a meeting of the Society for Pidgin and Creole Linguistics; to Randall Kim for permission to use this and other materials he distributed on Chinese Pidgin English; and to Phil Hubbard for comments on an earlier draft. My very special thanks to Colin Warner for his analysis of the transcript, and for researching the Hoover Library archives at Stanford University, where he found the transcript of *Rex vs. Kong*. In keeping with the citations, I retain the form 'vs.' rather than the currently more common 'v.' used in identifying legal cases.

contrast, inverted yes-no and *wh*-questions are more controlling. He concludes that although accommodation using intonation questions may result in more efficient court procedures, “when interpreted against background assumptions and expectations of formality and social distance some symptoms of intended friendliness and empathy may be taken to overaccommodate” (1991:127). In other words, since most defendants bring background assumptions and expectations about power asymmetries and formality to the trial situation, too much accommodation may be perceived in the same way that too much intimacy in the wrong situation is perceived: as image-threatening.

In this paper I examine a different kind of accommodation: the use of foreigner talk to defendants with limited English skills, in this case Chinese defendants. The primary data are from *People of the State of California vs. Ah Jake* (1887).<sup>2</sup> The transcript records the speech of (i) a Chinese defendant, Ah Jake, who is accused of murdering Wah Chuck, (ii) a witness, Ah Ting, who says he is a gardener who grows vegetables, (iii) an interpreter, Lo Kay, who interprets primarily for the witness, Ah Ting, and (iv) two lawyers: the judge and the district attorney. I will also refer briefly to another preliminary examination for a murder trial, *Rex (Crown of Canada) vs. Kong* (1914) for confirmation that the discourse practices identified are not unique to the *California vs. Ah Jake* trial.

What is striking from the point of view of legal discourse is that in both transcripts the lawyers use simplified language akin to foreigner talk in their questioning. Like the lawyers in Linell’s study, they use intonation questions and some non-inverted *wh*-questions. Such forms can be considered instances of choosing the less controlling form, but, given the context of other syntactic and lexical forms selected, they also appear to be features of foreigner talk. An example is provided in (1) from *California vs. Ah Jake*, which involves the judge (“Court”) first addressing the witness, Ah Ting, through an interpreter, and then addressing the defendant, Ah Jake, who interrupts at turn 5:

- (1) 1. Q. *Court*: [To Ah Ting] You see Wah Chuck touch Ah Jake; take hold Ah Jake?  
2. A. *Ah Ting*: (through interpreter) He say ah Jake lie. Wah Chuck no touch him.  
3. Q. *Court*: Wah Chuck no touch Ah Jake?  
4. A. *Ah Ting*: No.  
5. *Def*[Ah Jake]: Talk him [Ah Ting] lie. He help Wah Chuck; no like me—that two men kill me; he [Ah Ting] like him [Wah Chuck]; catch me.  
6. Q. *Court*: You say he take your money?  
7. A. *Def*: He take my \$30.  
8. Q. *Court*: One purse.  
9. A. *Def*: That bag.  
10. Q. *Court*: He take your purse; he take your pocketbook.

---

<sup>2</sup> The title page in the source-book, however, wrongly dates it as 1874.

## Courtroom Accommodation

11. A. *Def.* I put one bag that side; I put down here, gold, gold coin.  
12. Q. *Court.* He take all your money? (*California vs. Ah Jake* 51)

The judge uses partial repetition of a sentence in turn 3, presumably in part for purposes of confirmation and solicitation of clarification of the referent of *him* for the court record. Such uses of repetition complexify analysis of the judge's, and the district attorney's, foreigner talk, but there are enough examples of lawyers initiating non-repetitive turns in modified English (e.g. turn 6, 10, in which there is no tense marker), and turn 8 (use of *one* instead of *a*), to suggest they were not simply repeating the defendant's speech patterns. My hypothesis is that despite what are no doubt accommodation practices, the power semantic of linguistic disparity is nevertheless maintained by the use of foreigner talk.

### 1. The Linguistic Features

In order to understand this kind of discourse better, we need to look at the linguistic features used in more detail. In *California vs. Ah Jake* three main types of non-standard English are spoken:

- a) Pidgin English, spoken by Ah Jake, the defendant
- b) English as a second language, spoken variably by Lo Kay, the interpreter
- c) Foreigner talk English, spoken by the judge and the district attorney

I will outline the main features of each in §1.1.-1.3. The accuracy of the transcript is of course debatable. For one, it was written in shorthand: "I certify that the foregoing is a correct transcript from my shorthand notes" (57). For another, frequent mistakes have been noted in court transcripts. Furthermore, court reporters are known to attempt to change the speech of judges to make it appear more standard and less "crude and blundering" (Tiersma 1999:176, citing Walker 1990). However, as in most historical work, the text is all we have, and the usages appear consistent enough for it to be unlikely that there was an exceptionally high rate of editing by the court reporter in this case.

#### 1.1. Pidgin English

Ah Jake appears to speak Pidgin English. Although we might expect Chinese Pidgin English, one of the structures he uses, clitic object marking on the verb preceding an animate object (see characteristic (iv) below), does not appear to be a feature of this language; it is, however, used in pidgins spoken in New Guinea, Australia, and other parts of the Southwest Pacific (see Holm 1989). Ah Jake's pidgin includes absence of tense marking and:

- i) Negation expressed as pre-verbal *no*:
- (2)
- a. I *no* got money. (42)
  - b. I *no* can fight him. (50)

- ii) Zero copula (the following examples also illustrate his increasingly emphatic use of comparatives):
  - (3) a. He strong me. (50)
  - b. He stronger me. (50)
  - c. He man more strong me. (50)
- iii) Paratactic complex clause combining in semantically dependent constructions:
  - (4) a. I shoot Wah Chuck I go down here. (temporal, 56)
  - b. I no shootum him he shootum me. (conditional, 50)
  - c. He no hold my foot I no fall down at all. (conditional, 51)
- iv) Clitic object marking on the verb, as in (4b) and (5):
  - (5) a. Q. You want to ask him some questions?
  - A. Me askum him. (47)
  - b. He say killum me. (54)

## 1.2. English as a Second Language

The interpreter, Lo Kay, speaks what can be considered to be English as a second language (ESL), in the sense of non-native English probably learned without much or any schooling (see McArthur 1998). His speech is distinct from Ah Jake's pidgin in that it does not exhibit paratactic clause combining or object clitic marking. In many respects it resembles foreigner talk (see §1.3 below), except that Lo Kay is not modifying his native language. Like many ESL speakers, he shows both systematic second language features as well as variable control of target features (see Huebner 1985). (6) shows that he knows tense forms and *do*-support:

- (6) a. Q. Was he dead then?
- A. He was dead. (46)
- b. [reporting on an interruption by the defendant, Ah Jake]  
          He [Ah Jake] say don't know what he [Ah Ting] say. (44)

More often, however, we find absence of tense (see (7-9)), and of *do*-support,

- (7) [N]o see him dead that place. (45)

as well as absence of prepositions:

- (8) Wah Chuck he say Ah Jake steal his pair boot and sack rice. (43)

## *Courtroom Accommodation*

We also find uses of *one* for indefinite *a*:

- (9) He run down this way; give one piece of paper to one family house; he run very quick; then he write one piece paper to Lew Barnhardt. (44)

Occasionally we find absence of infinitive *to*, giving the impression of a serial verb construction, as in:

- (10) He say he run down get buggy come up take him down. (45)

The latter may be a feature of direct translation, however.

### **1.3. Foreigner Talk**

What then do the judge and the district attorney use? As expected, lawyerly language at times. At the beginning of *California vs. Ah Jake* the judge says to the defendant, who has claimed not to have enough money to hire an attorney,

- (11) Having no attorney I will say to you that it would be better for you not to say anything that will have a tendency to criminate you.<sup>3</sup> (42)

and at the end:

- (12) Order will be made requiring him to give bail in the sum of \$500 with two good and sufficient sureties to be approved by the Court. (57)

However, he does not use this register in questioning the witness or the defendant. Rather, he chooses accommodation devices. In the direct questioning portion the questions are short, and mostly in simplified English characterized by many features of foreigner talk (Ferguson 1975, see also Romaine 1988 and references therein):

- i) tenseless verbs
- ii) absence of *do*-support
- iii) preverbal negative *no*
- iv) absence of possessive clitic or preposition
- v) phrasal adverbs like *what for*

These are illustrated in (13):

---

<sup>3</sup> The Oxford English Dictionary cites several instances of this form from the mid-seventeenth century on.

- (13) a. Q. *Court*: How long you have that gold dust?  
A. *Def*: One month ago.  
Q. *Court*: What for you no sell him?  
A. *Def*: I keep buy grub. (52)
- b. A. *Def*: I see last night my coat burn. (shows his coat with holes in it) I under him—  
Q. *Court*: What you say the man name who gave you the \$30?  
A. *Def*: Ah Chung.  
Q. *Court*: What pocket you have \$30 in?  
A. *Def*: This. (showing left side)  
Q. *Court*: You have the other this pocket.  
A. *Def*: This pocket carpet sack.  
Q. *Court*: Little carpet sack this pocket (54)

Much of this foreigner talk appears to facilitate comprehension, although it can also be construed to be condescending (absent intonation this is somewhat hard to tell). But on one occasion it causes breakdown with the witness's interpreter:

- (14) *District Attorney*: I guess we will have to lead him a great deal.  
Q.<sup>4</sup> When Ah Jake shoot Wah Chuck, what Wah Chuck do?  
A. You mean Ah Jake shoot, what Wah Chuck do?  
Q. Yes.  
A. No say anything at all.  
Q. What he do after he get shot? After he get shot what he do?  
A. You mean Wah Chuck? (45)

Although the interchange becomes more functional it does not become fully so until a more standard style is used:

- (15) Q. He bleed any? Did he go back after he went to Barnhardt's? (45)

The district attorney then continues his questioning of the witness via the interpreter using tense and *do*-support most of the time, but maintaining non-inverted questions. He switches back to foreigner talk when Ah Jake takes the stand, and this time no breakdown appears to occur. Ah Jake, while readily admitting to the shooting, fails to establish that he was justified in doing so. There is no evidence that the lawyers' use of foreigner talk modifies the power relation between them.

## 2. The Second Transcript

We turn now to the second transcript. Here the defendant is a sixteen-year-old called Kong Yu John ("Jack Kong"), who, prior to the trial, had lived for four

---

<sup>4</sup> The transcript fails to indicate whether the question is posed by the judge or the district attorney at this point. I assume that when "Court" is not specified, the district attorney is speaking.

*Courtroom Accommodation*

years in the household of the Millards, and had some Canadian schooling. While halting, his speech is considerably more fluent than that of Ah Jake or the interpreter. The linguistic disparities are therefore not as large in *Rex vs. Kong* as they are in *California vs. Ah Jake*. However, similarities are evident. The defendant speaks ESL and the prosecutor uses foreigner talk. Jack Kong was accused of murdering Mrs. Millard, dismembering her, and burning her in a furnace. The prosecutor asks:

- (16) 1. Q. And they treated you very, very kindly?  
2. A. Yes.  
3. Q. Mrs. Millard treat you kindly?  
4. A. Yes.  
5. Q. Always?  
6. A. Well, she sometimes cross, very cranky.  
7. Q. Yes. Why?  
8. A. Being something which I had done not please her.  
9. Q. Being something which you have done not please her, would make her a little cranky with you? (*Rex vs. Kong* 6)

In this trial the defendant spoke very low, and so at times the lawyers appear to have been repeating for the benefit of the jury. This seems to be in part what the question in turn 9 is aimed at. We may note that the switch from Kong's *had* in turn 8 to *have* in turn 9 particularizes the "not doing" from habitual, iterative events to the day in question. This is actually not in the prosecutor's best interests, since he appears at this point to want to get Kong to admit to general frustration and anger with Mrs. Millard, not simply frustration on that day, seeing he has just asked at turn 5 in (16) about *always* and then goes on to ask in (17) about temperament. Whatever is going on in turn 8, the absence of *do*-support in turn 5, *Mrs. Millard treat you kindly?*, is not a case of repetition of the defendant's speech. Right after (16) the prosecutor asks,

- (17) Q. What was her temperament? Was she calm? (7)

and is stopped for asking a leading question. He tries again:

- (18) Q. Do you understand the word "temperament"?  
A. No.  
*Court:* Use another word, *disposition*.  
Q. What was her disposition?  
A. I do not understand that.  
Q. You do not understand that word. What was her nature? (7)

After this exchange, which again would benefit from information on intonation, there is a long sequence of turns in which the defendant's English is corrected, suggesting that *You do not understand that word* was used as a putdown. Exam-

ples appear in (19). (19a) corrects the lack of infinitive and copula verb, (19b) corrects the infinitive with the gerund, and (19c) corrects lack of tense marking:

- (19) a. A. You mean what caused her angry?  
Q. Yes, what caused her to be angry with you? (7)
- b. A. She insist me to do it.  
Q. She insisted on your doing it?  
A. Yes. (10)
- c. A. And then I carry her down to the basement.  
Q. You carried her down to the basement. Now how long between the time, how long between the time that you hit her with the chair and the time you carried her down the stairs, how long time elapsed? (12-13)

Again the defendant is not denying the murder (indeed, he had already confessed) but trying to give extenuating circumstances. Again the lawyers are not sympathetic. Rather, they seem to be using their exploration of the circumstances as a means for extracting further incriminating evidence.

### 3. Conclusion

This little study has shown that the power semantic can be present because of the structural properties of the situation, even when accommodating devices such as rising intonation questions and foreigner talk are used. At the same time, *California vs. Ah Jake* has shown us that the stereotype of the judge as upholder of form, linguistic as well as situational, does not always hold. We may have less stereotyped views of prosecutors, but here too *Rex vs. Kong* has shown us that style-shifting toward the defendant's speech is possible. Philips has discussed ways in which "[t]he ideological diversity enacted by the judges...is itself partly stimulated by and can be seen as an engagement with the defendants' resistance, showing how even resistance that does not articulate a clear ideological position can influence and shape the nature of ideological diversity in a local cultural scene" (Philips 1998:122). Philips is concerned with judges' political ideologies, and proposes that there is on the one hand a conservative-liberal continuum usually denied by the public, and on the other a record- vs. procedure-oriented ideology that is often not recognized at all; both of them play out in different forms of courtroom control. I propose that we add use of foreigner talk as well as non-native varieties as an additional window on to ideological diversity in interactional legal discourses.

### **Primary References**

- People of the State of California vs. Ah Jake. Superior Court of Sierra County, California, 23 Oct. 1887. In Albert Dressler (ed.), *California Chinese Chatter*, 42-57. San Francisco, 1927.
- Murder Case British Columbia. A Record of the Case of Rex (Crown) vs. Kong, in the Supreme Court of British Columbia, Vancouver, 18 May 1914. Hoover Institution Archives, Miner Chipmann collection, major Document BC40.

### **Secondary References**

- Brennan, Mark. 1994. Cross-Examining Children in Criminal Courts: Child Welfare under Attack. In Gibbons (ed.), 199-216.
- Brennan, Mark, and Roslin E. Brennan. 1988. *Strange Language: Child Victims under Cross Examination*. Wagga Wagga, Australia: Riverina Murray Institute of Higher Education.
- Conley, John M., and William M. O'Barr. 1998. *Just Words*. Chicago: University of Chicago Press.
- Danet, Brenda. 1985. Legal Discourse. In Teun van Dijk (ed.), *Handbook of Discourse Analysis* 1:273-291. London: Academic Press.
- Eades, Diana. 1994. A Case of Communicative Clash: Aboriginal English and the Legal System. In Gibbons (ed.), 234-264.
- Ferguson, Charles A. 1975. Towards a Characterization of English Foreigner Talk. *Anthropological Linguistics* 17:1-14.
- Gibbons, John, ed. 1994. *Language and the Law*. London: Longman.
- Gumperz, John J. 1982. Fact and Inference in Courtroom Testimony. In John J. Gumperz (ed.), *Language and Social Identity*. Cambridge, UK: Cambridge University Press, 163-195.
- Holm, John, ed. 1989. *Pidgins and Creoles*. Cambridge, UK: Cambridge University Press.
- Huebner, Thom. 1985. System and Variability in Interlanguage Syntax. *Language Learning* 35:141-163.
- Linell, Per. 1991. Accommodation in the Courtroom. In Howard Giles, Justine Coupland, and Nikolas Coupland (eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge, UK: Cambridge University Press, 103-130.
- McArthur, Tom. 1998. *The English Languages*. Cambridge, UK: Cambridge University Press.
- Philips, Susan U. 1998. *Ideology in the Language of Judges: How Judges Practice Law, Politics, and Courtroom Control*. New York: Oxford University Press.
- Romaine, Suzanne. 1988. *Pidgin and Creole Languages*. London: Longman.
- Shuy, Roger. 1993. *Language Crimes: The Use and Abuse of Language Evidence in the Courtroom*. Oxford: Blackwell.

Elizabeth Closs Traugott

Tiersma, Peter M. 1999. *Legal Language*. Chicago: University of Chicago Press.  
Walker, Anne Graffam. 1990. Language at Work in the Law: The Customs, Conventions, and Appellate Consequences of Court Reporting. In Judith N. Levi and Anne Graffam Walker (eds.), *Language in the Judicial Process*. New York: Plenum Press, 203-231.

Elizabeth Closs Traugott  
Department of Linguistics  
Margaret Jacks Hall, Building 460  
Stanford University  
Stanford, CA 94305-2150

[traugott@stanford.edu](mailto:traugott@stanford.edu)

## Discourse Pragmatics vs. Prescriptive Stylistics

OLGA T. YOKOYAMA

*University of California, Los Angeles*

A programmatic claim is made in this paper regarding the relationship between discourse pragmatics and stylistics. I suggest that the notion of prescriptive stylistics is ready to be reexamined in the light of the knowledge about language that has accumulated in recent decades, particularly its discourse-pragmatic, cognitive-psychological, and social aspects. Below I attempt to substantiate this claim empirically.

### **0. Introduction: Why Prescriptive Stylistics (PS)?**

Stylistics is a well-established area of language study, consisting of descriptive and prescriptive stylistics. Lucas (1955:16) seemed to grasp the dual nature of style when he defined it as two things: 1) a way of writing, and 2) a good way of writing. The first, descriptive stylistics, can be found in the functional treatments of the Prague School and of linguists such as Halliday, Crystal and Davy, as well as in literary and sociolinguistic treatments of topics such as register.

The second, prescriptive stylistics, is found primarily in handbooks. As a practical matter, prescriptive stylistics has a great deal of potential significance for and influence on non-linguists. It can be used to introduce students to linguistic concepts and can help with effective communication. Moreover, stylistics handbooks are a gold mine for discourse analysts, as they deal with areas of language that are not controlled by cut-and-dried grammatical rules but rather by elusive and hitherto ill-defined regularities and relations (unless, as is sometimes the case, the treatments are simply completely misguided). Despite its pervasiveness, however, prescriptive stylistics has been virtually ignored by linguists, with a few not very successful exceptions, e.g. Darbyshire (1971).

The recent tendency among stylisticians towards a lesser degree of normativeness nevertheless falls short of taking full advantage of advances in linguistics, and especially in discourse pragmatics. See, for example, Todd and Hancock's (1986:360-361) failure to distinguish between particles and prepositions in "This is the sort of behavior up with which I shall not put"; or Burchfield's (1996:577) mistaken claims about a double passive construction in "members who are found to have taken cocaine." More generally, however, as stated by Rannie (1915:130), a

champion of the study of stylistics in the early part of the 20th century: “Style is concerned not with correct or incorrect, but with good, better, and best.” Discourse-pragmatics, as an area with scalar rather than absolute distinctions, is well equipped to provide a *theoretical* foundation to the phenomena that have traditionally been treated in prescriptive stylistics.

## 1. Dangling modifiers and Voice

I will explore two of several syntactic constructions that are staples of English prescriptive stylistics: dangling participles and passives. In general, the complexities in the acceptability judgments of certain syntactic constructions have increasingly been recognized, hence a trend towards greater acceptance for at least some of them. Among formerly proscribed constructions that have virtually “made it” into the norm are split infinitives, whose rehabilitation seems to have begun with Jespersen and Curme, and which have steadily gained ground through Morris and Morris (1975), Strunk and White (1979), and Quirk et al. (1985:1121-1123), up to the latest edition of the Chicago Manual of Style (1993:76); and Close’s well-argued analysis (1987) was perhaps not without effect on their recent redemption. Lately, some authors prefer to allocate the use of some traditionally non-normative variants to colloquial and informal speech, and so refrain from condemning them indiscriminately (as in Todd and Hancock’s (1986) or Burchfield’s (1996) treatments of stranded prepositions).

### 1.1. Dangling modifiers (participles)

The treatments of dangling modifiers by stylisticians range from straight proscription (Morris and Morris 1975, Strunk and White 1979, Chicago Manual of Style 1993) to recommendations of avoidance in cases of potential ambiguity (New York Public Library Writer’s Guide to Style and Usage 1994, Todd and Hancock 1986). Interestingly, Burchfield (1996:805) acknowledges the actual rarity of ambiguity, but blames “the centuries-old failure to fault overt examples of unattached participles” in part for the present difficulty in making acceptability judgments. I will argue below that—on the contrary—rather than having anything to do with normativists’ permissiveness, dangling participles have managed to continue being (legitimately) generated *despite* the proscriptions of normativists.

Several types of dangling participles have enjoyed acceptance since the early part of the 20th century: a) absolute participles like “Strictly speaking, ...”, “Taking all things into consideration, ...”, “Putting it mildly, ...”; b) “generalized” participles like “When dining in the restaurant, a jacket and tie are required”; and c) in scientific publications, like “When treating the patient with.., the therapy consists ...” (Curme 1947, Quirk et al. 1985). Other types of dangling participles remain proscribed in English, although numerous literary attestations never fail to be listed, e.g. (1), an example from Jespersen (1964:94).

- (1) *He* felt himself gently touched on the shoulder, and looking around, *his* father stood before *him*. (Dickens).

Strunk and White (1979:13-14) recommend rewriting (2) as (3), and Quirk et al. (1985:1121-1123) suggest (7)-(9) for (4)-(6) respectively (the acceptability judgments for the latter are those of Quirk et al.).

- (2) Without a friend to counsel him, the temptation proved irresistible.
- (3) Without a friend to counsel him, he found the temptation irresistible.
- (4) ?Driving to Chicago that night, a sudden thought struck me.
- (5) ?Since leaving her, life has become pointless.
- (6) ?Walking down the boardwalk, a tall building came into view.
- (7) Driving to Chicago that night, I was struck by a sudden thought.
- (8) Since leaving her, I have felt that life was pointless.
- (9) Walking down the boardwalk, I saw a tall building.

It is important to note, however, that (3) and (7)-(9) are not exactly the same as (2) and (4)-(6). The distinction is particularly clear in (2) vs. (3): the main clause in (2) has a certain immediacy that (3) lacks. Narratologically, (2) would be a case of *style indirect libre* (Lips 1926), while (3) is the voice of an omniscient narrator. The same immediacy distinguishes (4)-(6) from (7)-(9), although it is impossible to ascribe it exactly to *style indirect libre*, given the first person narration. What unites (2) and (4)-(6) is the acuity of the psychological and perceptual experiences described in the main clauses: “The temptation is irresistible!”, “I know what!”, “Life is pointless!”, and “Wow! A tall building!”. This acuity is lost in (3) and (7)-(9), although these “corrected” versions may gain in normativeness. The difference in acuity, I would argue, amounts to a difference in point of view. In (2), the point of view is the character’s, whereas in (3) it is the narrator’s. Extending the same distinction to the other examples: in (4)-(6) the point of view is that of the speaker as s/he was at the time the narrated event took place, while (7)-(9) reflect the point of view of the speaker removed from the scene of the narrated event and placed in the speech event.<sup>1</sup> Pragmatically—and, of course, narratologically—the difference is significant.

There is another, not unrelated factor that accounts for the productive and regular generation of sentences with dangling participles, despite repeated proscriptions by stylisticians. Sentences like (2) and (4)-(6) are possible only when the deleted participial subject is topical. In narratives, this usually means being the topic of the preceding context. To that extent, subject deletions in dangling participles can be considered intersentential forward deletions, rather than intrasentential backward deletions. It is notable that Quirk et al. (1985) assume that the deleted participial subjects in (4)-(6) are all first person. Perhaps the relative prominence of “I” on the scale of topicality (cf. Kuno’s (1987) “Person Hierarchy”) has suggested such a reading of (4)-(6), although deleted 3rd person subjects would also be

---

<sup>1</sup> The terms “speech event” and “narrated event” are used here in the sense of Jakobson (1956:133 ff).

possible in (4)-(6), provided that the subjects are topical in the preceding context. Cognitively, the referential knowledge of the referent of the deleted subject must be assumable by the speaker to be among the knowledge items activated at the moment preceding the utterance of the preposed participial clauses;<sup>2</sup> the greater the justification for such an assumption, the more acceptable the sentence.

At least one of the conditions just suggested—the topicality of the deleted participial subject and the assumption in the main clause of the point of view of the participant of the narrated event—appears to be satisfied in all “good” sentences with dangling participles in English and in a number of other languages; consider the following examples:<sup>3</sup>

- (10) Hearing the floor creak behind me, my heart froze with fear, for I realized that Moriarty was inches away.
- (11) Zbole me glava, slušajući tu dreku.  
hurt me-acc head-nom hearing that fight-acc  
‘My head started to hurt, listening to that fight.’ (SC)
- (12) V takuju noč’, proxodja po cepjam, gusto mozgi nalivajutsja  
in such night passing through front-lines heavily brains swell  
dumami.  
thoughts-instr  
‘On such a night, while passing through the front lines, one’s brains heavily swell with thoughts.’ (R)
- (13) While driving through the snowstorm, dreading every curve, my car skidded helplessly over the icy road.
- (14) En chevauchant à travers la forêt, nos montures prirent peur.  
while riding across the forest our mounts became afraid  
‘Travelling through the forest on horseback, our mounts took fright.’ (F)
- (15) Bidding each other farewell, Holmes turned on his heels, while Watson proceeded in the direction of the morgue.
- (16) Polučiv èto izvestie, mnoju byli nemedlenno prinjaty  
receiving this notice by-me were immediately taken  
sootvetstvujuščie mery. (R)  
appropriate measures-nom  
‘Receiving the notice, appropriate measures were taken by me immediately.’
- (17) Upon awakening next morning, the somber knell of church bells reached our ears.
- (18) Having known Harry for five years, he still reminds me of a baboon.

<sup>2</sup> To paraphrase this in terms of knowledge sets, as discussed below in section 3: the referential knowledge of the referents of the deleted subjects must be in the set of current concern at the moment preceding the utterance; see Yokoyama (1986) for details.

<sup>3</sup> The following abbreviations are used: Cz = Czech, F = French, LV = Late Vedic Sanskrit, OR = Old Russian, R = Russian, SC = Serbian/Croatian.

- (19) Vracájući se uveče, dočekala me je mlaka crvenkasta svetlost.  
returning refl evening welcomed me past warm purple light  
'Returning home in the evening, a warm purple glow welcomed me.' (SC)
- (20) I ubdistasja ot sna, i poskočivši skoro ot lože svoego,  
and awaking from sleep and jumping quickly from bed self's  
napade na nix strax velik o videnii tom.  
fell on them fear great of vision that  
'And the two of them woke from their sleep, and jumping off their bed  
quickly, a great fear fell upon them about that vision.' (OR)
- (21) tám ha enam drstvá bhír viveda  
that-one ptcl him seeing fear found  
'And seeing him, fear overtook him.' (LV)
- (22) Vylezá z domu, takové nám byl mráz.  
going-out from house such to-us was frost  
'Going outside, it was so cold (on us).' (Cz dial.)

These examples contain cases of synecdochic antecedents in (10)-(12), split antecedents in (15), possessive antecedents in (13), (14), and (17), and oblique antecedents in (16) and in (18)-(21); in (22), the dative *nám* 'to-us', coreferential with the deleted participial subject, is not an argument but rather a so-called Dative of Interest, a strongly topical non-argument element that is frequent in Czech, especially spoken Czech (King 1998). All of these deletions are seriously problematic for a straightforward intrasentential syntactic analysis and must instead be accounted for by resorting to the discourse-pragmatic notions suggested above. Nor can straightforward prescriptive stylistics account for the high acceptability rating of these sentences even in languages in which the normative rule proscribing dangling participles is enforced, such as English, French, or Russian. Note, however, that no proscription exists in Old Russian, Serbian/Croatian, Sanskrit, or in Czech dialects. In modern languages with a strong prescriptive tradition and with a grammatical orientation towards sentential syntax, the formulation of the participial subject deletion rule has been couched in terms of subject identity, in all likelihood because in most cases the subject and topic of a sentence in these languages overlap. This is why Burchfield (1996) was misguided in ascribing the difficulty in judging the acceptability of dangling participles to a long-standing failure to fault these constructions: there was no reason to proscribe dangling participles in the first place, as long as they satisfied the two conditions suggested above. It is likely, rather, that native speakers' ambivalence towards them results from the longstanding but misguided proscription itself.

## **1.2. Voice**

Another example of an English stylistic rule that can benefit from a discourse-pragmatic perspective is the recommendation against the use of the passive voice. This recommendation stems from the claim that the passive voice is weak while the active voice is vigorous and direct. The influential figure Quiller-Couch (1923:120-

121) was quite forceful in his prescription: “Generally use transitive verbs, that strike their objects, and use them in the active voice, eschewing the stationary passive, with its little auxiliary *it*’s and *was*’s [...] by his use of the straight verb [...] you can tell a man’s style, if it be masculine or neuter”. Eastman (1978) and Strunk and White (1979) follow suit, advocating the “vigorous” and “forcible” active voice. Such judgments, evidently, arose from a cultural attitude that values “vigor” and “directness”, combined with a faulty meta-linguistic intuition that the grammatical *active* voice is *vigorous* while the *passive* voice is *weak*.<sup>4</sup>

To be sure, more nuanced considerations have also been voiced by grammarians. Notable among them are Jespersen’s claims (1964:120-121), which early on mention the effect of voice in shifting point of view: “the person or thing that is in the centre of interest at the moment is made the subject of the sentence”. (This, in fact, is an early formulation of what Kuno and Kaburaki proposed in 1977 as “Empathy”.) In (23), Jespersen said, “the greater interest” is “taken in the passive rather than in the active subject”:

(23) His son was run over by a car.

Quirk et al. (1985:943) also point to “the importance of the passive voice as the means of reversing the normal order of ‘agentive’ and ‘affected’ elements, and thus of adjusting clause structure to end-focus and end-weight”; a similar observation was made by Eastman (1978:153). This point, as well as Quirk et al.’s observation that “[T]he passive is generally more commonly used in informative than in imaginative writing, notably in the objective, non-personal style of scientific articles and news items” (808), are entirely consistent with the “Topic Hierarchy” and “Humaneness Hierarchy” proposed in Kuno (1987). The use of passive voice in the form of the “impersonal passive” has in fact been long accepted in PS; cf. Copperud (1964:289), who found “‘The issue was discussed for an hour...’ hardly objectionable if the discussers of the issue are of no moment.” Fowler, for his part, objected to those cases of impersonalization in which deleting the *by*-agentive deprives the hearer of information as to who the responsible party is, thus shirking the responsibility by using constructions like “It is felt that...” (1965:440).

Despite normativists’ recommendations to the contrary, the passive voice continues to be produced by speakers of English, and for a good reason. While Gross’s (1979) demonstration of the complexity of the passive voice clearly showed that it is premature to assume that its functions and conditions will be exhaustively delineated in the foreseeable future, it is hardly disputable today that the passive voice encodes at least one important pragmatic difference: it is used when the speaker assumes the patient’s point of view. Thus, I suggest that point of view is a category operative in both participial subject deletion and in passives. In fact, point of view is manifested in a number of disparate areas of language; cf.

---

<sup>4</sup> My Microsoft Word grammar check persistently suggests that I change all passive sentences into active ones—another example of prescriptive overgeneralization!

(24), which violates not only a point of view constraint on the subject selection of certain directional verbs like “come/go” (Kuno & Kaburaki 1977: 663-664) but also a point of view constraint on the lexical semantics of certain adverbs, such as “suddenly”:

(24) ??I suddenly came up to him.<sup>5</sup>

Point of view can explain the *raison d'être*, as well as the degrees of appropriateness, of these various constructions in their contexts. As such, I propose that it reflects a category of “theoretical stylistics,” a field that is yet to be created with the help of discourse pragmatics.

## 2. Russian word order and intonation

### 2.1. Word order and pragmatics

In so-called “free word-order languages”, word order is another area typified by scalar judgments, and, as is usually the case in such areas, it invites discourse-pragmatic considerations. Consider the following examples, all of which are discourse-initial and assume so-called “neutral” intonation:<sup>6</sup>

- (25) Ja rodilas' v Xarbine.  
I-nom born in X.  
'I was born in Harbin.'
- (26) Vam prislali cvety.  
you-dat they-sent flowers-acc  
'(Someone) has sent you flowers.'
- (27) Kolju izbili xuligany.  
K.-acc beat-up hooligans-nom  
'Kolya was beaten up by hooligans.'

Note that the sentence-initial elements in (25)-(27) are nominative subject, dative indirect object, and accusative direct object, respectively; in (26) there is no subject, although the verb is marked plural, and in (27) the nominative subject *xuligany* ‘hooligans’ is in sentence-final position. The sentence-initial placement of the first and second person pronouns, as I have argued elsewhere (Yokoyama 1986), is due to the normal presence of the referential knowledge of the interlocutors in their “sets of current concern” (the activated portion of the cognitive sets the interlocutors constitute). Under the conditions of normal interpersonal discourse, {I} and {you} are a given, unless displaced out of the sets of current concern in the course of a subsequent exchange.<sup>7</sup> This explanation, however, cannot account for

---

<sup>5</sup> The reading of this sentence improves if two selves are implied, as in a retelling of a dream.

<sup>6</sup> See Yokoyama (2001) for a discussion of Russian intonation.

<sup>7</sup> The capacity of a set of current concern is clearly limited, although its actual capacity cannot be determined without empirical data and experimentation. Note in this respect Spencer’s insightful remark that the addressee “has at each moment but a limited amount of mental power available.”

(27), which can be uttered without the presence of Kolya at the scene. What allows the order in (27) to be generated is the speaker's own concern for Kolya in combination with the assumption that the addressee, too, is concerned with Kolya. The speaker thus—perhaps presumptuously—posits the referential knowledge of {Kolya} into the addressee's set of current concern and proceeds with the ordering. This imposition of {Kolya} onto the addressee may or may not be reasonable, since Kolya may or may not be relevant to the addressee at the exact moment the utterance (27) is produced. It is thus a potential violation of relevance constraints, and as such it is likely to be produced, and accepted without the addressee's protest or confusion, between two very close individuals (like Kolya's parents), where the risk of unreasonable imposition is minimized. Were the same reality to be reported by Kolya's mother to the mother of one of the hooligans, the order would be exactly the same as in the corresponding English version "Your son beat up my son".

Russian PS remains silent on nearly all but the most obvious questions concerning word order. Yet word order is certainly an area where stylistic guidance could be used. The sentence-initial placement of referential items assumed by the speaker to be in the addressee's center of current concern is a matter of communicative competence. Notably, Lucas (1955:16ff.) stressed the importance of communication-based stylistics, actually proposing stylistic rules and recommendations strikingly similar to Gricean maxims more than a decade before Grice—another indication that stylistics and discourse-pragmatics overlap in their jurisdiction. That the word-order effect of the English passive voice, another discourse-pragmatic phenomenon, often corresponds to Russian word order variation (itself unaccompanied by a change of voice) is yet another piece of evidence that ties the word order changes in both languages to speakers' assessment of the cognitive stance of the addressee.

## 2.2. Word order, intonation, and pragmatics

Consider now the underlined items in (25)-(27). These items can be moved to the left, provided sentential stress is placed on them, and the items that appear after them in these new versions are both deaccented and lose their phrase boundaries. These versions differ from those in (25)-(27) in what is sometimes called tenor or register. I suggest elsewhere (e.g. Yokoyama 2001) that the difference in question concerns the fact that (25)-(27) belong to a distanced planned discourse, while the variations under consideration are spontaneous and are generated when the social/psychological interlocutor distance is short. In (25)-(27) the sentence-final position is reserved for knowledge items the speaker does not assume to be in the addressee's knowledge set, at least not in the propositional relationship of a given utterance. In planned discourse, this cognitive status is encoded linearly through sentence-final position, while in spontaneous discourse the encoding is intonational. Combined with the evidence that linear encoding is learned in the process of

---

Note also Miller's (1956) considerations of the human capacity for processing information (1917:3).

normative socialization (Yokoyama 2001), this last point suggests a relationship between Russian spontaneous discourse and Bernstein's "restricted code" (see e.g. Bernstein 1973).

There is a large body of literature on Russian spontaneous discourse as attested among the speakers of Standard Literary Russian (Zemskaja 1973, 1987, *inter alia*). The vernacular of the intelligentsia is distinguished from the formal standard by a cluster of linguistic and communicative features ranging from phonetics, intonation, phonology, morphology, word-formation, syntax, and lexicon to linguistic poetics and play, as well as conversational behavior. It is also in the vernacular that most genderlect features occur in Russian. The clustering of the features suggests a subsystem (which some Russian scholars call "colloquial style" and others "colloquial language"), or in fact two subsystems of educated Russian vernacular—a male and a female one.<sup>8</sup> Regardless of whether it is called a style or a register (or something else), in the case of the speakers of Standard Literary Russian the question of code switching becomes unavoidable.

### **3. Conclusion**

I have suggested above in section 2.1 that the speaker's assessment of the content of the addressee's knowledge set plays a role in the form of the utterance, down to the fine points of word order. The vernacular of the Russian intelligentsia, as outlined in section 2.2, also suggests that the speaker's assessment of his/her relationship with the addressee plays a role in the form of the utterance, including the fine points of sentential stress placement, intonational phrasing, and concurrent word order rules (not to mention a host of other phenomena). These are all linguistic choices the speaker makes, and they depend on extralinguistic, pragmatic factors which involve the speaker's (and the addressee's) cognitive (although not necessarily conscious) stance. Recalling Rannie's (1915) formulation that style is a matter of choice among good, better and best, these pragmatic factors, then, are categories in what has traditionally been called stylistics.

Point of view is another cognitive phenomenon that, as I have suggested in section 1, affects what has traditionally been called stylistic choices. The point of view shifts manifested in various linguistic phenomena testify to the human capacity not only to assume others' points of view (as is the case with the passive voice in section 1.2) but even to view one's own alter ego as a separate entity (as suggested above with respect to the participial constructions (4)-(6) vs. (7)-(9))<sup>9</sup> in section 1.1. The investigation of point of view can hardly be said to have been carried exhaustively in discourse linguistics; but once the relationship between the established field of stylistics and the newer field of discourse-pragmatics is accepted, one hopes that "[a] comprehension [...] of the general principles from which the rules [...] result will not only bring them home to us with greater force

---

<sup>8</sup> Cf. Yokoyama (1999b) for a description of the Russian genderlect system.

<sup>9</sup> Cf. also Yokoyama (1999a) for another manifestation of the "doubling" of the first person speaker in Russian possessive reflexives.

but will discover to us other rules of like origin” (Spencer 1917:3).

It was briefly suggested in section 2.2 that code switching is involved in the shift to the vernacular of the Russian intelligentsia. In fact, code switching is at play in any language where there is a choice among alternate systems of expression. The evidence examined here suggests that discourse-pragmatic mechanisms are employed not only in the choice of particular alternatives in each given utterance but also in more global switching between codes and registers.

The greater part of any English stylistics handbook is occupied by lexical choices, and a significant portion of lexical stylistics entries are concerned with historical changes in semantics and usage. These have not been considered in this paper. Many lexical choices that pertain to register, however, fit squarely into the same area as do choices concerning Russian vernacular code switching, for example. Longstanding recommendations to avoid foreign words (Fowler 1965:212, Strunk and White 1979:81), or to prefer “the Saxon words to the Romance” (Quiller-Couch 1923:120, also cf. Gowers 1948:60), are also a matter of assessing the addressee’s knowledge and the implications for the linguistic self-image of the speaker; both are discourse-pragmatic decisions.

To conclude, then, I have tried to show that discourse-pragmatics addresses itself largely to the same sort of phenomena as traditional stylistics. I have also suggested that the time may be ripe for discourse-pragmatics to attempt to provide a theoretical basis for traditional stylistics, as part of what may be called “Theoretical Stylistics”.

## References

- Bernstein, Basil. 1973. *Class, Codes and Control*. London-Boston: Routledge and Kegan Paul.
- Burchfield, R.W., ed. 1996. *The New Fowler’s Modern English Usage*. Oxford: Clarendon.
- Chicago Manual of Style, 14<sup>th</sup> edition*. Chicago: University of Chicago Press.
- Close, R. A. 1987. Notes on the split infinitive. *Journal of English Linguistics* 20: 217-229.
- Crystal, David and Derek Davy. 1969. *Investigating English Style*. Bloomington-London: Indiana University Press.
- Curme, George O. 1947. *English Grammar*. New York: Barnes & Noble.
- Darbyshire, A. E. 1971. *A Grammar of Style*. London: Andre Deutsch.
- Eastman, Richard. 1978. *Style: writing and reading as the discovery of outlook*. New York: Oxford University Press.
- Gowers, Sir Ernest. 1948. *Plain Words*. London: His Majesty’s Stationery Office.
- Gross, Maurice. 1979. The failure of generative grammar. *Language* 55: 859-885.
- Jakobson, Roman. 1957. Shifters, verbal categories, and the Russian verb. *Selected Works II*. The Hague-Paris-Berlin-Amsterdam: Mouton, 130-147.
- Jespersen, Otto. 1966. *Essentials of English Grammar*. University, AL: University of

*Discourse Pragmatics vs. Prescriptive Stylistics*

- Alabama Press.
- King, Katerina. 1998. *The Czech Dative of Interest: The Hierarchical Organization of Possession in Discourse and Pragmatics*. Ph.D. diss., Harvard University.
- Kuno, Susumu. 1987. *Functional Syntax*. Chicago-London: University of Chicago Press.
- Kuno, Susumu and Etsuko Kaburaki. 1977. Empathy and syntax. *Linguistic Inquiry* 8: 627-672.
- Lips, M. 1926. *Le style indirect libre*. Paris: Payot.
- Lucas, F.L. 1955. *Style*. London: Cassell.
- Miller, George A. 1956. The magical number seven, plus or minus two. *The Psychological Review* 63: 81-97.
- Morris, William and Mary. 1975. *Harper Dictionary of Contemporary Usage*. New York: Harper and Row.
- The New York Public Library Writer's Guide to Style and Usage*. 1994. New York: Harper Collins.
- Quiller-Couch, Arthur. 1923. *On the Art of Writing*. Cambridge: Cambridge University Press.
- Quirk, Randolph, et al. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.
- Rannie, David Watson. 1915. *The Elements of Style*. New York: Dutton.
- Spencer, Herbert. 1917 [1892]. *Philosophy of Style*. Boston-New York: Allyn and Bacon.
- Strunk, William Jr. and E. B. White. 1979 [1959]. *Elements of Style*, 3 ed. New York: Macmillan.
- Todd, Loreto and Ian Hancock. 1986. *International English Usage*. London- Sydney-Wolfeboro, NH: Croom Helm.
- Yokoyama, Olga. T. 1986. *Discourse and Word Order*. Amsterdam: Benjamins.
- Yokoyama, Olga. T. 1999a. The speech act empathy hierarchy and Russian possessives. In A. Kamio and K.-I. Takami (eds.), *Function and Structure*. Amsterdam-Philadelphia: Benjamins.
- Yokoyama, Olga. T. 1999b. Russian genderlects and referential expressions. *Language in Society* 28: 401-29.
- Yokoyama, Olga. T. 2001. Neutral and non-neutral intonation in Russian. *Der Welt der Slaven* XLVI: 1-26.
- Zemskaja, E. A. 1973. *Russkaja razgovornaja reč'*. Moskva: Nauka.
- Zemskaja, E. A. 1987. *Russkaja razgovornaja reč': lingvističeskij analiz i problemy obučenija*. Moskva: Russkij jazyk.

115 Kinsey Hall, Box 951502  
University of California, Los Angeles  
Los Angeles, California 90095-1502



**PARASESSION:  
GESTURE AND LANGUAGE**



# **Indicating Participants in Actions: From Prelinguistic Gestures to Signed Languages**

SHANNON CASEY

*University of California, San Diego*

## **0. Introduction**

In signed languages, verb agreement morphology consists of movement, displacement, or orientation of the hands in space to indicate agents, patients, recipients, sources, goals, or locations. Using evidence from prelinguistic gestures and verb agreement morphology produced by deaf children acquiring American Sign Language (ASL), I will argue that directional movement of the hands resembling verb agreement is not unique to signed languages but rather a basic gestural device for indicating participants in actions that has been incorporated and grammaticized into signed languages. Furthermore, I will suggest that grammaticization of directional movement began with its indexical use with present referents and was extended to absent referents.

## **1. Background: ASL verb directionality**

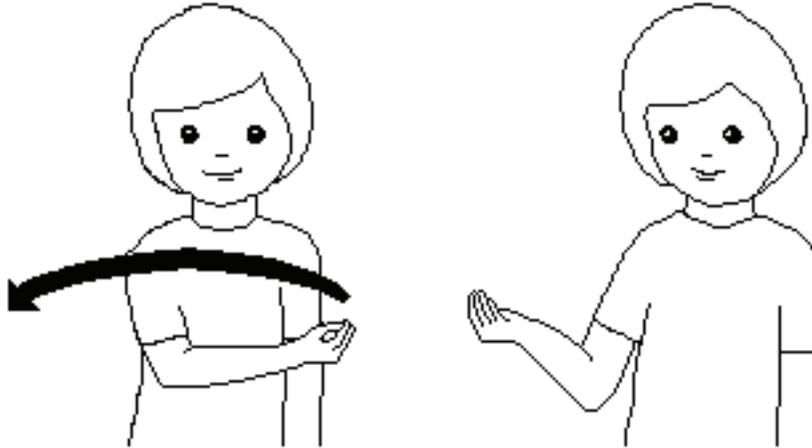
“Directionality” is defined as the use of movement, spatial displacement, and/or palm orientation in the production of a manual action gesture or sign to indicate an additional participant involved in the action. The term “participant” is used here to refer to an agent, patient, recipient, source, goal, or location. Verbs in ASL can show agreement with participants assuming these semantic roles through directionality (Fischer and Gough 1980; Padden 1988, 1990; Supalla 1982:59).<sup>1</sup> If participants are present in the immediate environment, the verb will show directionality with respect to the actual locations of the participants. If participants are absent from the immediate environment, places in space (called loci) can be established for reference to these participants (Friedman 1975:946). For example, if Sandy and Lee are not present, a locus for Sandy can be established on the

---

<sup>1</sup>Agreement in ASL is often described in terms of grammatical relations (e.g. subjects and objects). Semantic roles are used here (following the terminology of Goldin-Meadow and Mylander (1984) with the substitution of “agent” for their “actor”), because this study involves both signs and gestures. I do not want to imply that gestural productions are sentence-like through the use of grammatical relations.

signer's left and a locus for Lee can be established on the signer's right. To say that Sandy gives something to Lee, the sign GIVE moves from Sandy's locus (the agent) to Lee's locus (the recipient). See (1) below.

(1) SANDY-GIVE-LEE



An **index** refers to an object through a physical connection with that object and depends on the presence of the object for its interpretation (Peirce 1960, vol. 1:195-196, vol. 2:170). Thus, the indication of a participant present in the environment by movement or spatial displacement of a verb in ASL can be said to be indexical, because there is a physical connection between the spatial placement of the sign and the referent.

## 2. Current study

### 2.1. Method

Five deaf children acquiring ASL from their deaf parents were studied longitudinally between the ages of 0;08-2;11 (years;months). The data consisted of 37 videotaped free play and structured sessions with a range of 4-14 sessions for each child. The children's spontaneous action gestures and signs were analyzed for the use of directionality to indicate additional participants.<sup>2</sup> A reliability check between two coders (one of whom is a native Deaf signer) on a portion of the data showed 97.68% agreement on whether a production contained directionality.

Whether an action production was coded as a gesture or sign was determined by the following criteria in the order listed, from most to least important:

---

<sup>2</sup> The coding system used for these data was adapted mainly from that of Goldin-Meadow et al. (1995), including features from Liddell and Johnson (1989), Meier (1982), and T. Supalla (1982).

### *Indicating Participants in Actions*

1. form: What was the handshape, place of articulation, orientation, and movement of the production? Did the production resemble an ASL sign or a commonly occurring gesture, e.g. open handed 'give' or raised arms 'pick (me) up'? When examining the form of a production, Siedlecki and Bonvillian's (1993) finding that handshapes are the least accurate component of ASL signs produced by children aged 0;05-1;06 was taken into consideration.
2. semantic context: If the production looked like an ASL sign, did that meaning fit the context? If the production looked like a common gesture, did that meaning fit the context?
3. linguistic context: If the production followed an adult's utterance, it may have been an imitation of something the adult signed, and therefore a sign. If it was part of an utterance containing other signs, it may also have been a sign.
4. child's age: If the above criteria could not determine whether a production was a gesture or sign, the child's age was taken into consideration. Productions that could be clearly categorized indicated that children under age 1;11 produced more gestures than signs to refer to actions, whereas children aged 1;11 or above produced more signs.
5. native signer intuition: Productions that were still unclassifiable were shown to a native signer. If the signer could confidently determine whether a production was a gesture or sign, that judgment was coded.

A reliability check between the same two coders showed 98.98% agreement on whether a production was a gesture or sign. Although this agreement percentage is high, of the 521 gestures coded, 78% were 'give' and 'pick (me) up', which were often clearly gestures.

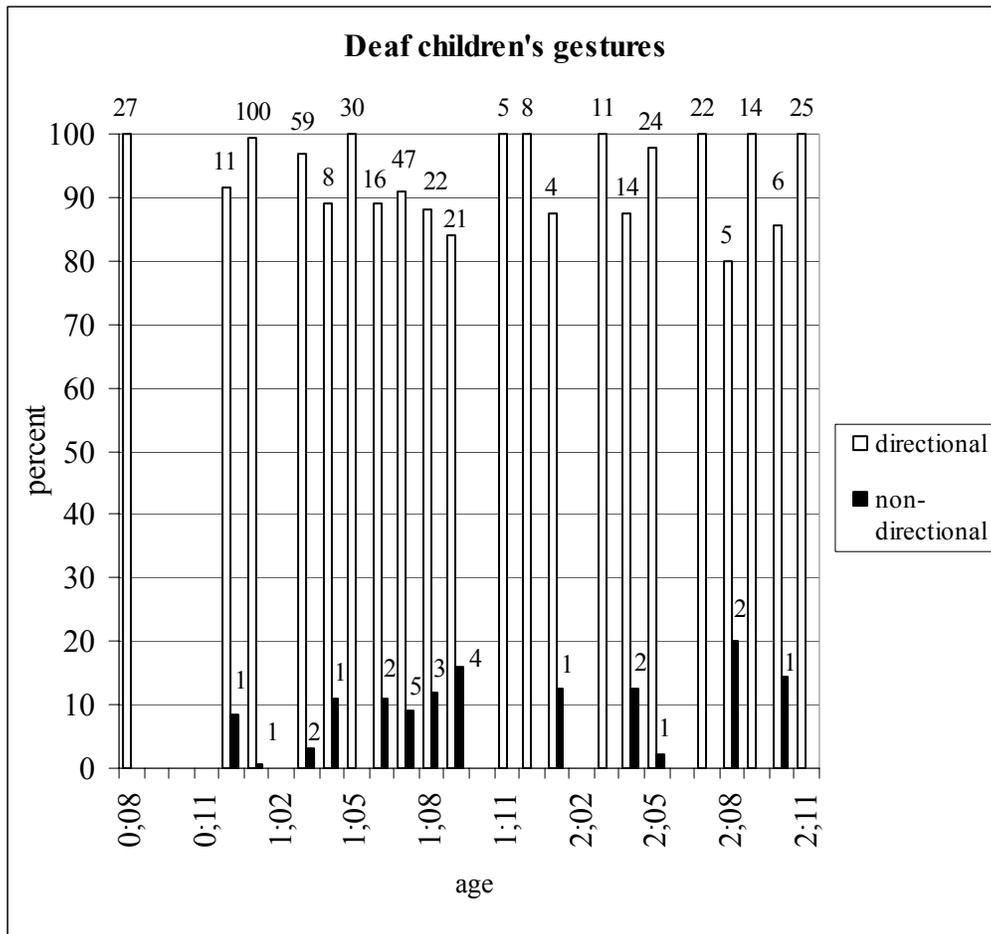
## **2.2. Directionality**

### **2.2.1. Gestures: Directional versus non-directional productions**

Deaf children were found to produce gestures containing directionality to indicate additional participants. Their spontaneous gestures were analyzed for the percentage of directional versus non-directional forms produced at each age. An example of a directional gesture is an open palm 'give' held out near a toy (the patient) to mean 'give toy'. In contrast, a non-directional gesture is an open palm 'give' that is not held near any participant. The use of directionality in gestures was productive in that the same gesture form could be used to indicate participants in various semantic roles. For example, Maggie (1;01) produced a 'give' gesture displaced near the agent ('you give') immediately followed by the same 'give' gesture displaced near the patient, i.e. a pile of toys ('give toys'), when the first gesture was not understood by the agent.

As shown in (2) below, most gesture productions were directional. A non-parametric rank F test for repeated measures<sup>3</sup> found the production of directional gestures (mean = 95%, mean rank = 12.32) to be significantly greater than that of non-directional gestures (mean = 5%, mean rank = 4.44):  $F_R(1,42) = 89.12, p < .0001$ .

(2) *Percentage of spontaneous directional versus non-directional gestures:*  
The numbers above the bars indicate the total productions at that age.



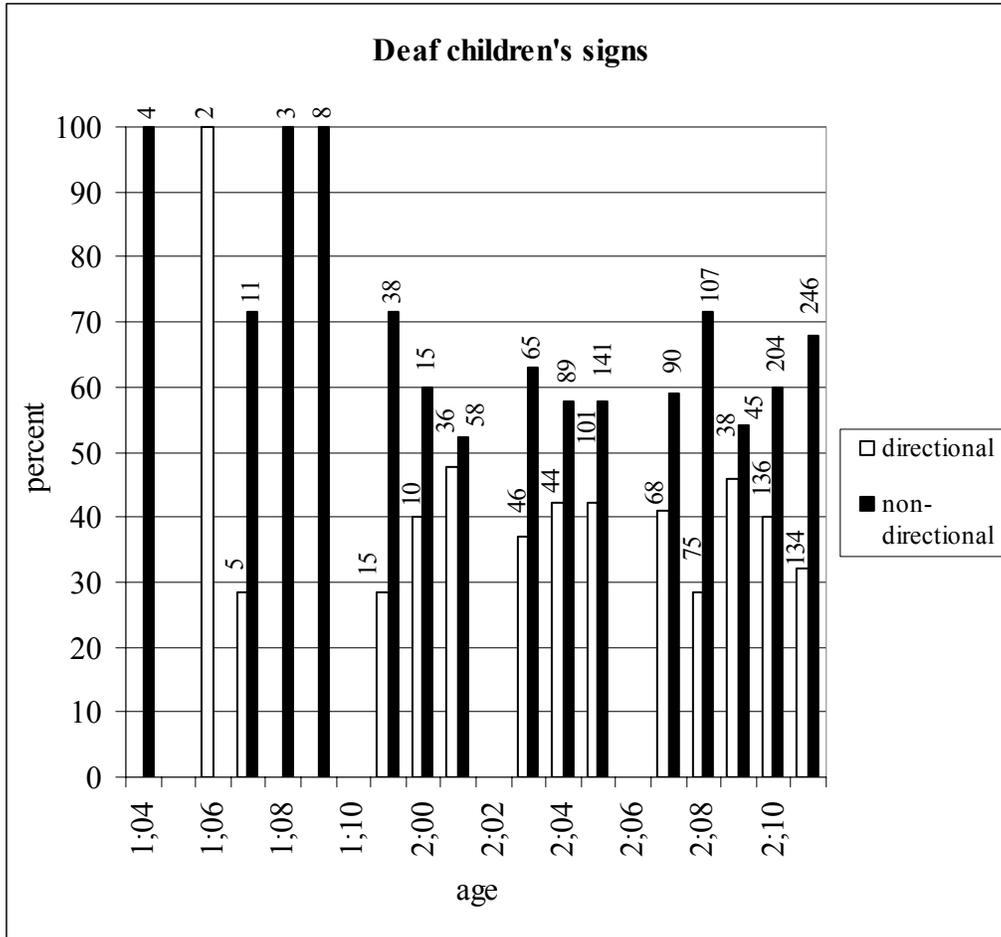
**2.2.2. Signs: Directional versus non-directional productions**

Deaf children's production of spontaneous ASL verb signs were analyzed for the presence of directionality at each age. For this analysis, ASL verbs that cannot occur with directionality and those produced with directional errors were excluded.

<sup>3</sup> A repeated measures ANOVA was not used because a modified Levene test found that the error terms did not have constant variance.

*Indicating Participants in Actions*

- (3) *Percentage of spontaneous directional versus non-directional signs:*  
The numbers above the bars indicate the total productions at that age.



As shown in (3) above,<sup>4</sup> verb signs were more often non-directional than directional. However, at most ages, at least 28% of signs were directional. The greater production of non-directional verb signs contrasts with the greater production of directional gestures. A nonparametric rank F test with repeated measures (cf. fn. 3) found the production of directional gestures (mean = 95%, ranked mean = 11.32) to be significantly greater than the production of directional signs (mean = 35%, ranked mean = 5.10):  $F_R(1,37) = 62.01, p < .0001$ .

This difference in the production of directionality with gestures versus signs may be accounted for by hypotheses concerning the late acquisition of ASL verb agreement morphology (Lillo-Martin 1991; Newport and Meier 1985). Children produced errors related to each of the following hypotheses proposed to account for this late acquisition of verb agreement morphology:

<sup>4</sup> These children did not produce any signs meeting the criteria for this analysis before age 1;04.

1. Not all verbs can occur with agreement (Lillo-Martin 1991:162; Newport and Meier 1985:931-932). Maggie (2;03) signed SLEEP toward a picture of sleeping dogs, however SLEEP cannot show agreement (for similar examples see Bellugi 1988:166-167; Fischer 1973).
2. Verbs differ with respect to the endpoint at which an argument can occur, e.g. some verbs move from the agent to the patient (HELP), whereas others move from the patient to the agent (HIRE) (Lillo-Martin 1991:162). To describe something falling on someone's head, Ben (2;04) signed FALL (with bent 5 classifier handshapes) starting on his head (the goal) and moving off. The movement of FALL should terminate at the goal, not originate at it.
3. Verbs differ in how many arguments they can agree with, i.e. some can agree with two arguments (GIVE), whereas others can only agree with one (WANT) (Lillo-Martin 1991:162). Patty (2;07) signed WANT moving from herself (the agent) toward a piece of paper (the patient) to say that she wanted the paper. WANT can be displaced near an agent (I-WANT) or patient (WANT-PAPER), but cannot traverse space to show agreement with two arguments in the same production (\*I-WANT-PAPER).
4. Verbs differ in the type of semantic roles with which they can agree, e.g. some verbs agree with agents and patients (HELP), whereas others agree with agents and recipients (GIVE) (Lillo-Martin 1991:162). Corinne (2;01) produced the sign FALL starting at her chin and moving off to mean 'fall on chin'. FALL can agree with the source and/or goal of the falling movement, but not with a body part of the intransitive agent (for similar examples see Fischer 1973; Meier 1982:122).
5. The spatial framework can shift (Lillo-Martin 1991:162). For example, a signer can take on the role of a third person by shifting the shoulders toward that referent's locus (Friedman 1975:950). In her description of a story involving a girl and a boy painting each other's faces, Maggie (2;11) ungrammatically used opposite sides of her face to stand for the girl and boy instead of moving her body to take on their roles. She signed PAINT (with an A classifier handshape) on the right side of her face to describe the girl being painted, and on the left side of her face to describe the boy being painted (for similar examples see van Hoek et al. 1987:118-120).
6. The production of pronoun or agreement reversal errors, e.g. use of the sign YOU to mean 'me'. Corinne (2;01) produced GIVE-YOU directed toward her mother to mean 'give me' when she was talking about her father giving her something, i.e. directionality with 'you' to mean 'me'.

Based on evidence from these directionality errors, I hypothesize that children produce less directionality with signs than with gestures because any type of directionality can be used with gestures, whereas directionality with signs is restricted by properties associated with specific ASL verbs. Thus, children may

opt to use non-directional forms more often due to constraints on the correct usage of directionality in ASL and their incomplete acquisition of these constraints.

The production of reversal errors has been found during the acquisition of pronouns in spoken languages (Charney 1980; Chait 1982; Clark 1978), and Petitto (1986) found pronoun reversal errors during the acquisition of ASL by deaf children between the ages of 1;09-1;11. The agreement reversal errors in the present study are particularly interesting in that this type of error was found in signs but not gestures. For example, the deaf children were not found to move 'give' gestures from a toy to the addressee (i.e. a gesture that seems to express the meaning 'give you toy') to mean 'give me toy'. The occurrence of reversal errors in signs, but not gestures, suggests that directionality has been grammaticized in ASL. If directionality in signed languages is like gestural pointing and contains no grammatical person distinctions, as claimed by Liddell (1995:26-27; 2000:315), the occurrence of reversal errors would not be expected in sign because they are not found in gesture. Gestural pointing is directed toward something to which someone intends to refer, whereas pronoun and agreement reversals are directed toward someone to whom the signer does not intend to refer. I suggest reversal errors occur because children have not learned grammaticized person distinctions in ASL, and that these errors are caused by conflicts between the primarily indexical use of directionality in gesture versus the grammatically constrained use of directionality in ASL.

### **3. Production of directionality when referents are present versus absent**

If referents are present in the immediate environment, children may be more likely to use directionality than if referents are absent, because present referents can be indicated indexically through directionality with the actual referent. On the other hand, to indicate absent referents, directionality must be produced with respect to one of the following:

1. the actual location of the object, but neither the object nor the location is visible: Maggie (1;07) produced a 'give' gesture displaced in the direction of a room that could not be seen to request that someone give her a blanket that was in the room.
2. a similar present object (Butcher et al. 1991:329): Ben (2;08) produced the sign POUR toward his own head while describing a story in which a girl and a boy poured water on each other's heads.
3. a habitual or previous location of the object (Butcher et al. 1991:329): Corinne (2;05) produced a 'give' gesture near a place that her sister had been sitting to mean that her sister would give her something.
4. a classifier or classifier-like use of a handshape to represent the spatial relationship between two entities: Maggie (2;11) produced the sign CLEAN with one hand and a classifier (bent 5 handshape) with the other hand representing holding a turnip to mean 'clean turnip'.

5. an arbitrary area in space (Butcher et al. 1991:329): Corinne (2;05) produced the sign LOOK-AT moving upward to the same area in space that she had previously used to comment on balloons to mean 'look at balloons'.

### **3.1. Gestures: Directionality with present versus absent referents**

When the referent was present, children produced an average of 95% of gesture endpoints with directionality, whereas when the referent was absent, they produced an average of 56% with directionality. However, a nonparametric rank F test with repeated measures (cf. fn. 3) did not find a significant difference between the production of directionality with present (mean rank = 4.87) versus absent (mean rank = 4.17) referents:  $F_R(1,11) = .0109, p = .9186$ .

The reason that a significant difference could not be found may be due to the small number of gestures produced with absent referents. Only three children produced gestures to indicate absent referents. Out of a total of 495 gesture endpoints, only 7 were produced when the referent was absent, 5 of which were directional. The two children who produced directional gestures with absent referents produced directional gestures with present referents at earlier ages.

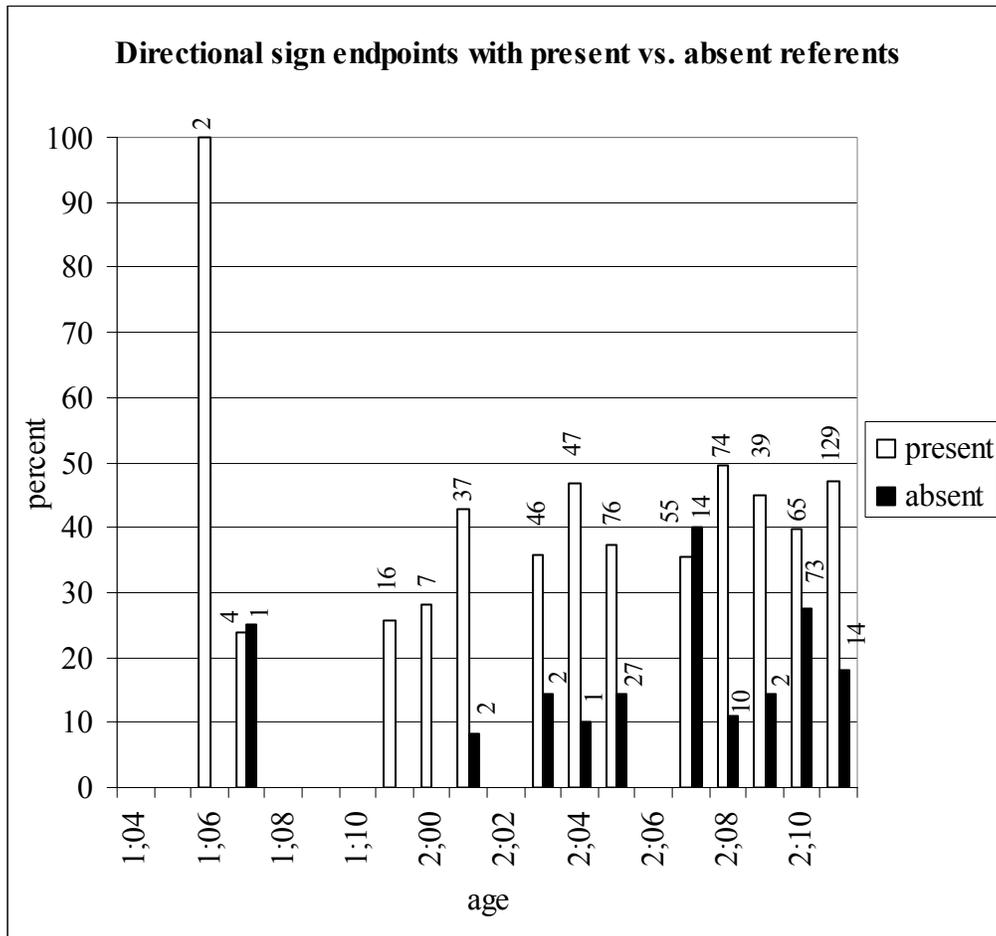
### **3.2. Signs: Directionality with present versus absent referents**

As shown in (4) below, children produced an overall higher percentage of directional sign endpoints when the referent was present than when it was absent. For this analysis, verbs that cannot occur with directionality and those containing the types of directional errors discussed in section 2.2.2 above were excluded. A nonparametric rank F test with repeated measures (cf. fn. 3) found that the production of directionality when referents were present (mean = 38%, ranked mean = 9.36) was significantly greater than the production of directionality when referents were absent (mean = 16%, ranked mean = 5.45):  $F_R(1,27) = 15.98, p = .0004$ .

Four of the five children produced directional signs to indicate absent referents. Three of these children produced directional signs to indicate present referents at earlier sessions than their first production of directional signs to indicate absent referents. The fourth child produced her first directional signs to indicate present and absent referents at the same age, i.e. 1;07. These findings suggest that it may be easier for children to direct signs toward physically present referents than to direct signs toward something standing for an absent referent.

*Indicating Participants in Actions*

- (4) *Percentage of directional sign endpoints with present versus absent referents: The numbers above the bars indicate the total productions at that age.*



**4. Similarities and differences in directionality with gestures and signs**

Directionality is similar in gestures and signs in that it is used for the same purpose and it is productive. Directionality is used in both gestures and signs to indicate additional participants involved in an action. In gestures, directionality is productive in that the same gesture form can be used with directionality to indicate various semantic roles. Productivity in signs is indicated by children's directionality errors, e.g. the use of directionality with ASL signs that cannot occur with directionality, and the use of directionality to indicate semantic roles with which particular verbs cannot agree. Another similarity between gestures and signs is that some directionality errors in ASL signs resemble directionality use in gestures. For example, Maggie (1;11) produced the sign GIVE held stationary near some stickers (the patient) to mean 'give stickers'. The sign GIVE can traverse space between an agent and a recipient or a patient and a recipient, but it

cannot be held stationary in space. On the other hand, 87-95% (varying by child) of 'give' gestures were stationary and displaced near a participant, and 57-89% of 'give' gestures were stationary and displaced near the patient in particular. Thus, this directionality error of a GIVE sign held near a patient resembles the most common form of directionality with 'give' gestures.

Directionality differences in gestures and signs include a greater use of directionality with gestures compared to signs and a lack of reversal errors in gestures. Furthermore, there is a difference in the age at which directionality was produced with gestures versus signs. The production of directionality in gestures was found at younger ages than in signs for four of the children. The fifth child was found to produce directionality in both gestures and signs in her first session (1;07), but this child was the oldest at her first session.

Directionality use with present referents is greater and earlier in both gestures and signs. The production of directionality was greater with present, as opposed to absent referents, although this difference was found to be statistically significant only for signs. With one exception, children who used directionality to indicate present and absent referents, produced directionality with present referents at a younger age than with absent referents in both gestures and signs.

The similarities in directionality use with gestures and signs, together with the fact that directionality occurs earlier in gestures, suggest that directionality use during sign acquisition is based on its earlier use in gesture. Furthermore, the finding that directionality with present referents is produced earlier and more often suggests that indexical indication of referents is easier for children than indication of a referent that is absent from the immediate environment.

## **5. Conclusion**

Based on similarities in directionality use with gestures and signs, I propose that directionality is a basic gestural phenomenon. Similarities in form, purpose, and productivity suggest that directionality in gestures and signs is related. Thus, directionality may have gestural origins not only in the acquisition of signed languages but also in their historical development. However, the occurrence of directional reversal errors in signs implies that directionality has been grammaticized in signed languages and is no longer purely gestural.

The production of directionality with present versus absent referents also has implications for the origins of signed languages. Children's greater production of directionality when referents are present and their earlier use of directionality with present referents suggest that they had difficulty indicating absent referents through directionality. Similarly, in her study of the emergence of Nicaraguan Sign Language, Senghas (1995:149) found that some older, less-skilled signers brought outsiders into a conversation to represent absent referents, because they had trouble using directionality with abstract places in space. These findings indicate that the grammaticization of directionality during the development of signed languages began with indexical indication of referents present in the environment and was later extended to absent referents.

### **Acknowledgements**

I wish to thank Ursula Bellugi for giving me access to videotapes that supplied data for this study. I also thank Susan Goldin-Meadow and Carolyn Mylander for training me in gesture coding. I am grateful to Joy Spurlin for coding and judgments, and to Helsa Borinstein, Melissa Herzig, Amy Hoshina, and Shane Marsh for native signer judgments. I am indebted to Edith Casey for Figure 1. I am very appreciative of comments provided by Karen Emmorey, Robert Kluender, Carol Padden, and participants at presentations related to this paper. All shortcomings and errors are entirely my own.

### **References**

- Bellugi, Ursula. 1988. The acquisition of a spatial language. In F. K. Kessel, ed., *The development of language and language researchers: Essays in honor of Roger Brown*, 153-185. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Butcher, Cynthia, Carolyn Mylander, and Susan Goldin-Meadow. 1991. Displaced communication in a self-styled gesture system: Pointing at the nonpresent. *Cognitive Development* 6: 315-342.
- Charney, Rosalind. 1980. Speech roles and the development of personal pronouns. *Journal of Child Language* 7(3): 509-528.
- Chiat, Shulamuth. 1982. If I were you and you were me: The analysis of pronouns in a pronoun-reversing child. *Journal of Child Language* 9: 359-379.
- Clark, Eve V. 1978. From gesture to word: On the natural history of deixis in language acquisition. In J. Bruner and A. Garton, eds., *Human growth and development: Wolfson College lectures 1976*, 85-120. Oxford: Oxford University Press.
- Fischer, Susan. 1973. The deaf child's acquisition of verb inflection in American Sign Language. Paper presented at the Linguistic Society of America Winter Meeting, San Diego.
- Fischer, Susan and Bonnie Gough. 1980. Verbs in American Sign Language. In W. Stokoe, ed., *Sign and culture: A reader for students of American Sign Language*, 149-179. Silver Spring, MD: Linstok Press.
- Friedman, Lynn A. 1975. Space, time, and person reference in American Sign Language. *Language* 51(4): 940-961.
- Goldin-Meadow, Susan and Carolyn Mylander. 1984. Gestural communication in deaf children: The effects and noneffects of parental input on early language development. *Monographs of the Society for Research in Child Development* 49(3-4): 1-151.
- Goldin-Meadow, Susan, Carolyn Mylander, and Cynthia Butcher. 1995. The resilience of combinatorial structure at the word level: Morphology in self-styled gesture systems. *Cognition* 56: 195-262.

- Liddell, Scott K. 1995. Real, surrogate, and token space: Grammatical consequences in ASL. In K. Emmorey and J. Reilly, eds., *Language, gesture, and space*, 19-41. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Liddell, Scott K. 2000. Indicating verbs and pronouns: Pointing away from agreement. In K. Emmorey and H. Lane, eds., *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, 303-320. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Liddell, Scott K. and Robert E Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies* 64: 195-277.
- Lillo-Martin, Diane. 1991. *Universal grammar and American Sign Language: Setting the null argument parameters*. Dordrecht: Kluwer Academic Publishers.
- Meier, Richard P. 1982. Icons, analogues, and morphemes: The acquisition of verb agreement in American Sign Language. Ph.D. diss., University of California, San Diego.
- Newport, Elissa and Richard P. Meier. 1985. The acquisition of American Sign Language. In D. I. Slobin, ed., *The crosslinguistic study of language acquisition, vol. 1: The data*, 881-938. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Padden, Carol A. 1988. *Interaction of morphology and syntax in American Sign Language*. New York: Garland Publishing, Inc.
- Padden, Carol A. 1990. The relation between space and grammar in ASL verb morphology. In C. Lucas, ed., *Sign language research: Theoretical issues*, 118-132. Washington, D.C.: Gallaudet University Press.
- Peirce, Charles Saunders. 1960. *Collected papers. vols. 1 and 2*, Charles Hartshorne and Paul Weiss, eds. Cambridge, MA: Harvard University Press.
- Petitto, Laura Ann. 1986. *From gesture to symbol: The relationship between form and meaning in the acquisition of personal pronouns in American Sign Language*. Bloomington, Indiana: Indiana University Linguistics Club.
- Senghas, Ann. 1995. Children's contribution to the birth of Nicaraguan Sign Language. Ph.D. diss., MIT.
- Siedlecki, Theodore and John D. Bonvillian. 1993. Location, handshape, and movement: Young children's acquisition of the formational aspects of American Sign Language. *Sign Language Studies* 78: 31-52.
- Supalla, Ted. 1982. Structure and acquisition of verbs of motion and location in American Sign Language. Ph.D. diss., University of California, San Diego.
- van Hoek, Karen, Lucinda O'Grady and Ursula Bellugi. 1987. Morphological innovation in the acquisition of American Sign Language. In *Papers and reports of child language development*, vol. 26, 116-123. Stanford, CA: Leland Stanford Junior University.

casey@ling.ucsd.edu

## Co-expressivity of Speech and Gesture: Manner of Motion in Spanish, English, and Chinese\*

SUSAN DUNCAN

*University of Chicago and National Yang Ming University, Taiwan*

### 0. Abstract

Languages such as Spanish and English differ in how each lexically packages and syntactically distributes semantic content related to motion event expression (Talmy 1985, 1991). Comparisons of spoken Spanish and English (Slobin 1996, 1998) reveal less expression of manner of motion in Spanish. This leads to the conclusion that ‘thinking for speaking’ in Spanish involves less conceptualization of manner. Here we assess speech-associated thinking about manner on a broader basis by examining not only speech but also the speech-synchronous gestures of Spanish, English, and Chinese speakers for content related to manner of motion. Speakers of all three languages produce manner-expressive gestures similar in type and frequency. Thus, motion event description may in fact involve conceptualization of manner to roughly the same extent in all three languages. Examination of gesture-speech temporal synchrony shows that Spanish manner gestures associate with expression of the ground component of motion in speech.

We consider these findings in relation to two assertions: (1) gesture compensates for content speech lacks, (2) gesture and speech ‘jointly highlight’ shared or congruent semantic content. A compensation interpretation of the Spanish manner gestures raises questions about the role of gesture data in studies of thinking-for-speaking, generally. Further evidence from a follow-up study, in which narrators had no visual exposure to the cartoon, lead us to interpret Spanish speakers’ manner-expressive gestures as an instance of joint highlighting. This interpretation accords with McNeill’s (1992) “rule of semantic synchrony” between speech and gesture, one of the foundations of his ‘growth point’ theory of language production (McNeill 1992; McNeill and Duncan 2000). We discuss some impli-

---

\* The research was supported by a grant from the Spencer Foundation of Chicago. Preparation of the paper was supported by a grant from the National Science Foundation KDI program, Grant No. BCS-9980054. The following people collaborated on much of what is described in this paper: David McNeill, Lisa Miotto, Juan Pablo Mora, Karl-Erik McCullough, Nobuhiro Furuyama, Asli Özyürek, Glenda Miranda, Hui-fang Hong, and Lily Wong.

cations of a joint highlighting interpretation for analyses of thinking for speaking and for lexical semantic theory.

### **1. Introduction and overview**

Speakers of all languages gesture spontaneously when they converse, tell stories, or narrate events. Many of these gestures manifest semantic content; for instance, they may iconically represent actions and entities that speakers refer to in narrative discourse. What interests many researchers about gestures is their potential as a communicative resource. Indeed, a principle of *compensation* is often applicable to the gesture-speech relationship (Taub and Piñar, this volume). Gestures may communicate propositional, semantic, and other content for which the resources of the spoken language are inadequate (Kendon 2000). There is little disagreement that gestures have a communicative function. However, gesture is an exceedingly heterogeneous domain of behavior. A principle of compensation by itself falls short of explaining all of the patterning we observe.

Compensation, for instance, does not explain something that characterizes a great deal of the gesturing that occurs in narrative discourse; namely, the significant degree of ‘overlap’ in the meanings expressed by gestures and speech during the intervals in which they synchronize. Much gesture production occurs in tight synchrony with speech, as it is linked to the prosodic and rhythmic structure of language (Tuite 1993). Components of motion such as path, manner, figure, and ground are often simultaneously expressed in the two modalities. For instance, if an English speaker describes a cat climbing up and a spontaneously-produced gesture synchronizes with her speech, the gesture is very likely to be some kind of iconic depiction of a figure climbing upward. Her two hands may represent the cat’s paws. They may move alternately as the cat’s paws would in climbing. They may also move generally upward in front of her body. Certain aspects of the relationship between the two modalities, we see, are better characterized as *joint highlighting* of semantic content, rather than as compensation by one modality for the expressive limitations of the other.

Some studies have focused on this tendency of gesture to track the semantic content of speech. Such work is in accord with a theoretical claim that, in each unit of language production in connected narrative discourse, co-produced speech and gesture are two simultaneous aspects of a single idea unit (McNeill 1985, 1992). We claim this unity despite obvious and significant differences between the two modes of expression, in terms of the way each structures meaning. Gesture is often imagistic, analog, and synthetic, while the speech ‘code’ is analytic, consisting of a linear-sequential arrangement of arbitrary units that derive from paradigms of categorial oppositions. McNeill has proposed a “rule of semantic synchrony” to capture the generality of the phenomenon of overlapping meanings in co-produced speech and gesture (McNeill 1992:28). Some cross-linguistic comparative research indicates that speakers of different languages gesture differently about events they describe, in ways that are semantically

aligned with how their languages constrain them to speak about those events (Kita 2000; Mueller 1994; McNeill and Duncan 2000).

Any psychological theory of speaking that attempts to deal with the fact of speech-synchronous gesture must acknowledge and account for the significant semantic overlap between the two modalities. Here we will propose also that lexical semantic theories of meaning can profit from careful consideration of exactly how the semantics of the two modalities link up.

## **2. The present study**

The analyses of speech and gesture reported here draw on the linguistic typological framework established in Talmy's cross-language comparisons of motion event expression (1985, 1991) and elaborated in Slobin's (1987, 1995, 1996, 1998) comparisons of the spoken and written forms of a number of languages, including Spanish and English. Talmy distinguishes 'verb-framed' and 'satellite-framed' languages, of which Spanish and English, respectively, are instances. The typology concerns which sentential constituent typically contains a lexical item expressive of the path component of motion. In Spanish sentential main verbs are typically path-expressive. *Entrar* 'enter' and *subir* 'ascend' are examples. In English, main verbs are more often motion- and/or manner-expressive, with path of motion expressed in satellites to the verb; for example, *go up*, *climb out*, *come down*, and *run in*.

Slobin (1998:7) notes that English has a "huge lexicon" of manner-expressive verbs compared to Spanish; further, that in spoken Spanish there are grammatical barriers, of sorts, to the use of manner verbs. In many Spanish discourse contexts it is cumbersome to include expression of manner of motion, as it must typically be incorporated using adjuncts to the verb. Such additions result in more elaborated constructions than may be rhetorically suitable in many discourse contexts. Slobin (1996, 1998) notes that manner, therefore, is frequently left out of Spanish motion event descriptions. On the basis of such evidence, Slobin builds a case for the existence of a cross-linguistically variable cognitive process of 'thinking for speaking.' This is the process by which thinking is structured into forms appropriate for spoken language production. Slobin claims that the dearth of manner-expression in spoken Spanish is evidence that thinking for speaking in that language involves less conceptualization of manner than it does in English. He claims further that experience with speaking Spanish over a life time results in a tendency among these speakers to attend less to manner of motion in the world, it being less "codable" in that language (1995, 1998:4).

These analyses have been based solely on speech and its written forms. Analysis of speech-synchronous gesture suggests itself as a method for broadening the examination of thinking for speaking. Attempts to assess the cross-linguistic variability of this process, that undertake to examine only speech and its written forms, carry a risk of circular reasoning.

A skeptical view [...] could hold that these differences operate only at the level of linguistic expression. To counter such a view, some way is needed to externalize cognition in addition to language. [...] we consider speech and gesture jointly as an ‘enhanced window’ onto thinking and show how the co-occurrences of speech and gesture in different languages enable us to infer thinking-for-speaking in Slobin’s sense. (McNeill and Duncan 2000:142)

We examine videotaped, unrehearsed, cartoon story narrations collected from adult Spanish, English, and Mandarin Chinese speakers; more than 100 narrations in all. Each of the speakers watched a 6.5-minute, animated cartoon about a cat and a bird and was then audio-videotaped telling the story of the cartoon to a naïve listener. We sample descriptions of events in the cartoon eliciting stimulus that involve a figure moving along a path in relation to a ground. In each of these, a particular manner of motion is an element of the depicted event. With one exception, the analyses we report below are limited to descriptions of just two of the motion events depicted in the cartoon eliciting stimulus. In one, the cat climbs up a drainpipe on a building. In the other, the cat rolls down a hill after having swallowed a large bowling ball.

We will describe a within-language (Spanish) difference in the way these two events are described, related to use of manner-expressive lexical items. Note that the ‘climbing’ and ‘rolling’ events each comprises the same set of motion event components—figure, path, manner, and ground. On the basis of Talmy’s typology, therefore, we have no *a priori* basis on which to predict within-language differences in the lexical-semantic and syntactic resources speakers will use to fashion descriptions of these events. We also observe cross-language differences in gesture form and gesture-speech synchrony. We will claim that these differences are informative concerning Spanish thinking for speaking.

### 3. Descriptions of rolling down a hill

The speech-synchronous gestures of the English, Spanish, and Chinese speakers, below, are representative of the range of gestural forms we encounter in our narration data. These excerpts also illustrate the consistency with which similar semantic content appears simultaneously in the two modalities. In (1), an English speaker performs a two-handed gesture synchronized with the spoken phrase *rolls on down*, in which the hands trace vertical circles alternately while moving along a path slightly downward and away from her body.

- (1) **he rolls on down** into a bowling alley
  
- (2) va **rodando** va rodando y entra a un establecimiento de boliche  
goes **rolling** goes rolling and enters to an establishment of bowling  
he goes **rolling**, rolling and enters a bowling place

- (3) ta jiu **gun gun gun gun** gun gang-hao gun-jin yi-jia bao-ling qiu guan  
he then **roll roll roll roll** roll just-good roll-enter a-CL bowling ball place  
then he **rolls and rolls and rolls**, then rolls into a bowling alley

In (2), a Spanish speaker, in exact synchrony with *rodando* ‘rolling,’ traces a path rightward with his right hand in front of his body. The hand flaps up and down as it moves, depicting path and a rolling manner. In (3), a Chinese speaker moves both hands from left to right and downward, the fingers wiggling. This path-manner gesture synchronizes with the reduplicated manner verb *gun* ‘roll.’

These are all examples of a type of multi-directional gestural motion we observe when narrators describe cartoon events depicted as having manner of a repeating, alternating, or agitated kind; for example, climbing, flying, and tip-toeing. In each case, the gesturing body part moves repetitively in more than one direction. The movement may be revolving, up and down, side to side, or multi-directional in an erratic pattern. In our analyses, ‘agitated’ gestural motions such as these are interpreted as expressions of manner of motion.

Two issues concerning the narration excerpts above will be relevant to our cross-language comparisons. The first concerns the temporal synchrony of speech and gesture. The speech in each example that is rendered in bold face represents the interval during which the ‘stroke’ phase of the co-occurring gesture is executed. A gesture stroke is the phase within the overall gestural movement during which the gesture’s semantic content is interpretable (see Kendon 1980:212; McNeill 1992:25).<sup>1</sup> The three excerpts are representative of what we observe across the three languages in regard to speech-gesture synchrony. When manner-expressive iconic gesture strokes co-occur with utterances containing manner-expressive lexical items, the two expressions of manner typically synchronize, often quite precisely.

The second issue is that the spoken portion of the Spanish excerpt, similarly to the English and Chinese, includes a manner-expressive lexical item, *rodando* ‘rolling.’ Manner-expressive terms are frequent in the Spanish speakers’ descriptions of this cartoon event. Twenty-two, or 68%, of thirty-two Spanish speakers who recalled and narrated this event included one or more manner-expressive lexical items in their descriptions of it; thirty-two, 86%, of thirty-seven English speakers did so. The most common grammatical construction used among the English speakers is the phrase *goes rolling*; among Spanish speakers, *va rodando* ‘goes rolling’, an analogous construction. We also see other manner verbs; for instance, *bounce* in English, *botar* ‘bounce’ in Spanish. Thus, despite an 18% difference between the speaker groups in use of manner verbs here, the spoken

---

<sup>1</sup> We examine the synchrony of stroke phases with speech closely on videotape with the aid of professional-grade VCRs. Such equipment makes it possible, in slow-motion mode, to listen to the audio portion of the recording while watching the video. The ability to slow the media down to frame-by-frame viewing speed without loss of access to the audio track is essential for the type of analyses we report here.

descriptions of the ‘rolling’ motion event are not dissimilar. For comparison, Slobin (1998:6) reports in regard to one motion event in the ‘frog story’ narrations that involved manner of motion, none of the Spanish speakers’ verbs was manner-expressive, while 32% of those used by English speakers were.

#### 4. Descriptions of climbing up a pipe

##### 4.1. Speech

The utterances below illustrate typical verb use in spoken descriptions of our other target motion event. These descriptions illustrate how the three languages form something of a continuum with respect to the tendency to use manner-expressive verbs. This fact will prove useful, presently, when we consider the hypothesis that gestural manner compensates for manner missing from speech.

- (4) Sylvester climbs up the drainpipe gets to the top
- (5) Silvestre empezó a subir por un tubo de desagüe  
sylvester he-began to to-ascend via a pipe of drainage  
sylvester begins to ascend via a drainpipe
- (6) mao kai-shi cong shui-guan pa pa-pa-pa pa-shang-qu  
cat begin via drainpipe climb climb-climb-climb climb-up-go  
the cat starts to climb the drainpipe he climbs and climbs he climbs up

English speakers often use the manner-expressive verb *climb* in one or more of the phrases they put together to describe the cat’s (“Sylvester”) ascent. Often a phrase like this combines in sequence with one or more additional phrases incorporating a non-manner-expressive verb, as in (4), above. In addition to the verb *get*, other non-manner-expressive verbs that occur in the English descriptions of this event include the verbs *go* and *come*. Neither is it unusual for English speakers to use *go* and *come* with no manner adjunct in their descriptions of the ‘rolling’ event. On the basis of descriptions of these two events, as well as others in our English sample, we can say that English offers its speakers a fair degree of flexibility to choose non-manner-expressive verbs to describe all or part of both the ‘climbing’ and the ‘rolling’ motion events in the cartoon. Examples include, *as he’s going up*, and, *he comes out the bottom of the drainpipe*.

In contrast, the Spanish descriptions of the ‘climbing’ event are almost without exception organized around path verbs. In this, these descriptions are more representative of our Spanish narration data in its entirety than were these speakers’ descriptions of the ‘rolling’ event. Use of *trepar* ‘to climb’ to describe this event occurs only twice in almost 40 Spanish narrations. This in spite of the fact that the cat climbs up the pipe several times during the cartoon, providing Spanish speakers with ample opportunity to use that manner-expressive verb, were they so inclined. For this event, Spanish speakers prefer path-expressive

*subir* ‘to ascend.’ *Entrar* ‘enter’ and *meter* ‘enter’/‘insert’ also occur. Therefore, here, our Spanish narration data are quite in accord with Slobin’s (1998) findings.

Chinese is among the satellite-framed languages according to Talmy’s typology (Talmy 1985; cf. Slobin and Hoiting 1994:102). In our Chinese-language data, path-expressive verbs and deictic verbs are rarely deployed as the main verbs of motion-event descriptive utterances. Further, as (3), above, and (8), below, illustrate, there is often repetition of manner verbs within utterances in ways not attested in the Spanish and English data. As a consequence, there sometimes seems almost a super-abundance of manner coloration in Chinese motion event descriptions.<sup>2</sup> The utility of the Chinese narration data for our comparison of speech and gesture has to do with these speakers’ very heavy use of manner verbs, compared to Spanish and English speakers. In this respect, Chinese is at the opposite end of a continuum from Spanish, with English in between. The language has many manner verbs and appears to make the use of them almost obligatory in narrative discourse contexts like this one. This difference will prove useful, presently, when we examine the gesture data for evidence in support of the gesture-speech compensation hypothesis.

#### **4.2. Gestures that express ‘climbing’ manner**

Now we examine the three speaker groups’ manner gestures in the context of descriptions of the ‘climbing’ target motion event. The English and Chinese descriptions, (7) and (8), provide further illustration of the phenomenon of gesture-speech semantic synchrony. Observe how precisely the motion event component-expressive gestures synchronize with speech expressive of the same components, both manner and path.<sup>3</sup> In (7.1), this English speaker’s right hand bumps up and down, an ‘agitated’ movement expressive of manner, while moving upward. In (7.2) the right hand changes to an index point up and continues, minus the bumping, on the same upward path, now moving through the speaker’s left hand, which is in the shape of the letter *C*, an iconic representation of the drainpipe.

---

<sup>2</sup> We note two features of our Chinese discourse data that may relate to the abundance of manner-expression. Example (6) illustrates both (a) the prominent manipulation of aspectual distinctions—incipient, progressive, and perfective—that leads to repetition of manner verbs within event descriptions, and (b) the very common use of directional verb complement (DVC) constructions, which relegate path-expressive and deictic verbs to a subordinate role in utterance-final position. There appear to be some limitations on the use of the latter verbs outside of DVCs.

<sup>3</sup> We employ the following typographic conventions for annotating the narration transcripts. Asterisks (\*) indicate self-interrupted speech; forward slashes, brief silent pauses; angle-bracketed text (<...>), filled pauses or syllable lengthening; curly-bracketed dots {...}, unintelligible syllables. Square brackets define individual gesture movements against co-occurring speech. Gestures can ‘nest’ and this is indicated with nesting square brackets. Gesture stroke phases are in bold face. Underlining represents intervals of temporary cessation of gestural movement, in-place, called gesture ‘holds.’

- (7.1)<sup>4</sup> [[ / so he **starts climbi**][ng] {manner+path}
- (7.2) / **through** the rainpipe / ]] {figure/ground+ground+path}
- (8.1) [ / **pa shui-guan-de wa**][[**i-bian** {manner+path+figure}  
 [ / **climb drainpipe-POS out**][[**side**  
 [ / **climbs the drainpipe's out**][**side**
- (8.2) **pa pa pa** {manner+path}  
**climb climb climb**  
**climbs and climbs**
- (8.3) [pa-**shang**]]-qu {path only}  
 [climb-**up**]]-go  
 [climbs **up**

In (8.1) the Chinese speaker's two hands flap alternately, moving up. The left hand continues on by itself in (8.2). Note that these two manner gestures occupy a long interval, during which the manner verb *pa* 'climb' is uttered four times. (8.3) illustrates how gesture can 'surgically' target an element of speech. The bold face shows that the path-only gesture stroke—an index point moving up—skips the manner-expressive verb to synchronize with the path-expressive, *shang* 'up.'

In the Spanish, we see a different situation. In (9), as with descriptions of the 'rolling' event, a Spanish speaker produces manner gestures in association with her motion event description. The difference here, of course, is that the manner gestures now synchronize with utterances that contain no manner verbs. Two path-expressive verbs are used to describe how the cat gets to the bird's window via the drainpipe. The speaker gestures non-stop through this spoken sequence. This is not unusual when speakers are fully engaged in the narration task.

- (9.1) entonces busca la ma][nera /\_][de **entra**][r {ground}{path+ground}  
 and then he-look-for the ma][**nner** /\_][of to-**ent**][er  
 and then he looks for the w][**ay** /\_][ / to **ent**][er
- (9.2) / **se met**][[e {manner+figure+path}  
 / **REFL he-ent**][[er  
 / **he ent**][[ers

---

<sup>4</sup> Starting from this point, the semantic content of the speech-synchronous gesture, as components of motion, will be indicated in curly brackets at the right margin. The notation 'figure/ground' indicates that the gesture in some way makes the relationship between the figure and ground components explicit.

*Co-expressivity of Speech and Gesture*

- (9.3) **por el** [desagüe/] {manner+figure/ground+path+ground}  
**via the** [drainpipe/]  
**via the** [drainpipe/]
- (9.4) [ / sí? ] {ground+path}  
[ / yes? ]  
[ / yes? ]
- (9.5) [desagüe entra / ] {manner+figure/ground+path+ground}  
[drainpipe he-enter / ]  
[drainpipe enters / ]

We will discuss four issues that emerge from consideration of this Spanish narration excerpt. First, the speaker uses no manner-expressive verbs or verb adjuncts. Second, four of her seven gestures are manner-expressive. Third, all the occurrences of manner in gesture temporally synchronize with an interval of speech that is somehow expressive of the ground component of motion (*desagüe* ‘drainpipe’), or of the relationship of ground to figure and path. The two path-expressive verbs, *entrar* ‘enter’ and *meter* ‘enter’/‘insert,’ for example, encode a path/figure-ground relationship, as does the preposition *por* ‘via.’ Fourth, this speaker’s manner gestures are of two types, one of which is a gestural form unattested in the other speaker groups.<sup>5</sup> We see the form in the first bracketed gesture of (9.3). The left hand curves in the shape of the pipe, the ground component. The index finger of the loose right hand points up as the hand executes a repeating, spiraling, ‘corkscrew’ motion upward. The second gesture in (9.3) repeats the performance, as does (9.5). In (9.2) the speaker flaps her hands alternately, moving them up, representing path and manner of motion.

This is quite a lot of gestural manner, especially considering the absence of manner verbs in the accompanying speech. In the next section we compare Spanish with English and Chinese in terms of overall frequency of manner gestures. We ask whether manner gestures of the sort produced by the Spanish speaker in this case function generally to compensate for the lower use of manner verbs in Spanish. Following that, we focus on the linkage, in gestured and spoken expression, between the manner and ground components of motion in Spanish.

---

<sup>5</sup> This sequence has been discussed previously in print (McNeill 2000; McNeill and Duncan 2000). Here we provide further details, in order to highlight the linkage between gestured manner and spoken ground, also the link between manner and ground in the gestures themselves. This one speaker’s manner gestures have been described as a diffuse ‘fog’ (McNeill 1997) blanketing the sequence. That description may obscure the features that interest us in this case; namely, the rather precise articulation of the two types of manner-expressive gestures in relation to ground-expressive speech; also, in relation to the non-manner-expressive gestures with which they are interpolated.

### 5. Overall frequency of manner in gesture, all three languages

Earlier, we described a continuum relating the three languages, in terms of relative amounts of manner expression in speech. The narration excerpts given so far demonstrate that some gestural expression of manner is a factor in all three languages. However, we might imagine that Chinese speakers produce fewer such gestures overall, since they do not really ‘need’ to gesture about manner of motion, it being abundant in their speech. A gesture-speech compensation claim is supported if manner-expressive gestures are most frequent in the Spanish narrations, least frequent in the Chinese narrations, with English somewhere in between. Counts of the manner-expressive gestures that speakers of the three languages produce with their motion event descriptions are presented in Table 1.<sup>6</sup>

**Table 1. Manner gestures: Overall counts and the content of co-occurring speech.** (20 speakers each; Spanish, English, Chinese)

| SPEECH                       | SPANISH   | ENGLISH   | CHINESE   |
|------------------------------|-----------|-----------|-----------|
| ground / figure / path       | 42        | 14        | 15        |
| motion (+path / ground)      | 9         | 5         | 1         |
| manner                       | 13        | 13        | 30        |
| manner + path / ground       | 7         | 24        | 30        |
| other                        | 17        | 2         | 12        |
| <b>total manner gestures</b> | <b>88</b> | <b>58</b> | <b>88</b> |

Spanish and Chinese speakers produce the same number of manner gestures. Each group of twenty speakers produced a total of 88. The twenty English speakers produced a total of 58. That Chinese and Spanish speakers both gesture quite a lot about manner of motion is inconsistent with the compensation claim, as stated in terms of relative amounts of manner expression in accompanying speech.

If compensation is not a factor, then we might expect the number of manner gestures to be about the same across the three groups, yet the count for English is just 66% that of the others. The most likely source of this disparity is the differing prosodic structures of the three languages, as this relates to how ‘idea units’ are packaged for production. As Tuite (1993) has noted (see also Kendon 1980; Nobe 1996), gesture stroke production reflects not so much the clausal as the prosodic structure of a language. The significant factor in occurrence of gesture strokes is

---

<sup>6</sup> These counts draw on a slightly more comprehensive sampling of *five* target motion events, rather than just the two that are largely the focus of this paper. These events are (1) cat runs across street, (2) cat climbs up on outside of drainpipe, (3) bird flies out of cage, (4) cat climbs up inside drainpipe, (5) cat and ball roll down hill. The sample was limited to twenty cartoon narrations from each speaker group. We selected from among the Spanish and Chinese speakers those who were least proficient in English as a second language. Note also that, if the counts of manner gestures overall seem low, given the size of the speaker sample, be aware that speakers do not reliably recall every cartoon motion event, sometimes do not gesture at all, or, of course, gesture about various other components of motion.

the patterning of intonation and pauses within individual utterances; specifically, gesture stroke phases tend to track peak prosodic emphasis. Inspection of individual speech-gesture productions in the three languages reveals that the difference in gesture counts, above, is likely at least in part attributable to the difference between English and Spanish utterances illustrated in (10) and (11).

(10) the cat **comes rolling** out {manner+path}

(11) [sale **rodando**] [ / **el gato**] {manner+path}  
 [he-exit **rolling**] [ / **the cat**]  
 [he exits **rolling**] [ / **the cat**]

Examples (6) and (8) from the Chinese data, cited earlier, reveal a pattern that contrasts with English in a similar way. A comparison of rates of gesture across these languages properly takes the motion event as its unit of analysis. Such a comparison reveals a higher ratio of gestures per unit, overall, in Chinese and Spanish than in English. While it is true that there are two manner gestures in (11) to just one in (10), in both cases the expression of manner is relevant to a single proposition expressive of one motion event. On this metric, gestural manner may be considered to be roughly equal across the three languages.

## 6. The link between manner and ground components of motion

### 6.1. Direct synchrony

Table 2 presents the percentages of manner-expressive gestures in the three languages that synchronize with manner-expressive speech *versus* non-manner-expressive speech. These percentages reflect the data presented in Table 1, collapsed to highlight the manner/non-manner distinction.

**Table 2. Manner-expressive gestures: Percent occurrence with ground, manner, or ‘other’ in speech. (20 speakers of each language)**

| SPEECH                      | SPANISH | ENGLISH | CHINESE |
|-----------------------------|---------|---------|---------|
| manner (+path/ground)       | 23 %    | 64 %    | 68 %    |
| ground/figure/path (motion) | 58 %    | 33 %    | 18 %    |
| other                       | 19 %    | 3 %     | 14 %    |

The shaded portions of the table reveal that, in Spanish, the largest portion of manner-expressive gestures synchronize with ground-expressive speech, while in both English and Chinese, the largest portion of these gestures synchronize with speech containing a manner verb or adverbial.

We can cite many examples of the pairing of ground-expressive speech with manner-expressive gestures in Spanish. When this occurs in the context of the ‘climbing’ event, often the gesture has the upward spiraling ‘corkscrew’ form



suggestive of a ‘semantic synchrony’ between manner and ground in Spanish thinking for speaking. Given the gestural evidence that a great deal of speaking-related conceptualization of manner occurs in Spanish, these findings of co-produced representations of ground and manner in speech and gesture are the basis for a claim that manner conceptualization in Spanish thinking for speaking builds on the ground component of motion. Seen in terms of the analytic structure of spoken expression, the components of manner and ground, certainly seem distinct. The semiotic modality of gesture, however, lets us see them as less separate in certain contexts, more as two aspects of one kind of motion.

## **6.2. Sequential links between manner and ground**

Examples like (13) are sometimes observed when we analyze motion event expression unfold across a connected sequence of utterances.

- (13.1) o sea [se su][be {path only}  
or it-is [REFL he-asc][end  
or it’s he asc][ends
- (13.2) **por un**][**tubo** / ][{...} es como un **desagüe** / ] {ground}  
**via a**] [**pipe** / ][{...} it-is like a **drainpipe** / ]  
**via a**] [**pipe** / ][{...} it’s like a **drainpipe** / ]
- (13.3) [ / **y se sube por allí** {path+manner}  
[ / **and REFL he-ascend via there**  
[ / and he ascends via there]

In between two path-focused spoken descriptions, the speaker produces three speech-gesture combinations expressive of the ground component, the drainpipe. The gestures in (13.2) are all iconic representations of the pipe. The gesture that synchronizes with the speech in (13.1) is expressive only of the upward path of motion. It is an index finger path trace. The gesture that synchronizes with (13.3) is a modified repeat of (13.1). Again, the index point moves up, however this time the gesture incorporates the ‘corkscrew’ motion, expressive of manner. Here we see how conceptualization of manner can also build sequentially on ground.

## **7. A follow-up study of Spanish speakers and listeners<sup>7</sup>**

Some have hypothesized that the manner gestures of Spanish narrators are direct gestural manifestations of the visual imagery these speakers retain from watching the cartoon stimulus. This is a testable alternative to our proposal that the gestures instead reflect a variety of thinking for speaking in which conceptualization of manner is linked to the component of ground. The cartoon does indeed contain

---

<sup>7</sup> Juan Pablo Mora, PhD, of the University of Seville, Spain, collaborated in running this follow-up study and in interpreting the results.

many depictions of exaggerated manners of motion. We conducted a follow-up study with eight native speakers of Spanish to test the visual imagery hypothesis.

The eliciting stimulus was the audio track (only) of one of our main sample of Spanish narrations. We selected a narration that was typical in terms of how the narrator described the two target motion events of ‘rolling’ and ‘climbing’. The ‘rolling’ event description incorporated *va rodando* ‘goes rolling’; the ‘climbing’ event, *sube por un tubo* ‘ascend via a pipe’. The latter also incorporated three utterances descriptive of the static setting, the drainpipe. Four of the eight participants listened twice to the narration on earphones, then told the cartoon story to one of the other four native Spanish speakers.

Thus, the narrators in this instance neither saw the cartoon itself, nor viewed another Spanish narrator speaking and gesturing about it. Our interest, of course, was in whether the narrators in this case would produce manner gestures at all; further, whether the gestures would exhibit the same forms and synchrony with speech as those of the speakers in the main study. Very briefly, analysis of the speech and speech-synchronous gestures revealed a difference only in overall amount of gesture between the two groups; otherwise, the follow-up study participants’ performances were similar in speech and in production of manner-expressive gestures. Their gestures exhibited the same range of forms, including the ‘corkscrew’, and similar speech-gesture temporal relationships.

## 8. Discussion

Slobin (1998) assesses Spanish speakers’ intuitions concerning the manner-related mental imagery they retain from reading narrative texts in Spanish. Speakers report that they retain very little. Such data are counter evidence to the claims we make here on the basis of speech-synchronous gesture. The gesture data, however, indicate that Spanish speakers do engage in manner-related thinking during acts of motion event description in the type of narration task we use. Our overall results provide reasons to question how severe the constraints may be on conceptualization of manner in Spanish. The results of the follow-up study also call into question any claim that the manner gestures of the speakers in the main study are due solely to having visually processed the action-packed cartoon. These results in fact comport with our proposal concerning the nature of conceptualization of manner in Spanish narrative discourse; that is, that it links in some way to the Spanish speaker’s conception of the setting or ground with respect to which a figure’s motion occurs.

That there may exist an intrinsic relationship between manner and ground should not be an entirely alien notion to English speakers. The relationship is lexically encoded in English manner verbs such as *dodge*, *slog*, *trek*, and *slide*. The overall tendency of English, however, may be to associate manner with figure (McNeill 2000). Neither, we would say, is this proposal of a manner-ground linkage in Spanish at odds with the contrasts between Spanish and English, considered in their entirety, that Slobin has reported (1995, 1996). Specifically, he notes that, where the grammatical and lexical resources of English promote

expression of manner and complex paths, the same resources in Spanish promote elaboration of reference to settings. This is where our gesture data and our proposed linkage between manner and ground seem relevant to linguistic theories of meaning. In elaborating on the ground or settings in which a particular motion event take place, the Spanish speaker simultaneously builds a conception of the manner of motion that must be involved. This is how we explain the within-Spanish-language difference between descriptions of the events of ‘rolling’ and ‘climbing’. The cartoon depiction of ‘rolling’ includes elements of the cartoonist’s imaginative distortion of reality, resulting in a stimulus event meeting Slobin’s criteria for an “exceptional manner” (1998:6), likely to elicit a manner-expressive verb from Spanish speakers. In our terms, the difficulty for the Spanish speaker is that there is no linguistically-specifiable figure-ground relationship from which the cartoon cat’s manner of motion could be the emergent result.

A theoretical issue is whether to interpret the fact of speech-synchronous manner gestures in Spanish as informative concerning the Spanish thinking for speaking process itself; indeed, of gesture generally as informative concerning the nature, generally, of thinking for speaking. To assert that gesture *is* informative is in keeping with the theoretical and methodological rationales that motivate much current cross-linguistic research in this area; namely that, “gesture and its synchronized speech express the same underlying idea unit. [...] By looking at the speech *and* the gesture jointly, we are able to infer characteristics of this underlying idea unit that may not be obvious from speech alone” (McNeill and Duncan 2000:143). In this formulation, image processing is hypothesized to be inseparable from processing of the spoken linguistic unit(s). The alternative is to interpret gesture as instead a manifestation of some other visuo-spatial cognitive process. Given the facts of speech-gesture synchronization, this would have to be a process that runs concurrently with thinking for speaking about motion events, but is somehow separate from it. The notion that gesture results from a process such as this accords with Hadar and Butterworth’s (1992) claim that gesture comes from visual imagery via a “direct route”. Gesture is, in their words, “the motor manifestation of imagistic activation”.

A weakness of formulations that rely on a ‘direct visual route’ and a semantic compensatory function of gesture is their inability to explain why the temporal co-occurrence of productions in the two modalities is non-random and so precise. A proposal that there is some kind of semantic ‘synchrony’, or congruence, between manner and ground, has the virtue of acknowledging the data on temporal synchrony. Nor can ‘direct visual route’ or compensation explain the occurrence of a Spanish-specific gestural form, the ‘corkscrew’. In general, these formulations are limited in their ability to account for systematic, cross-linguistic differences in gesture timing and gesture forms. For that matter, systematic *within*-language differences in form and timing are also a challenge (Duncan 2002). These limits flow from the fact that these formulations have, in important respects, little to work with beyond human hands and human vision. Consider the ‘corkscrew’ gestural form. It is problematic to adopt the position that the English or Chinese

speakers in our sample have limitations on their ability to move their hands in this way. It would be similarly problematic to claim that Spanish speakers' visual perception of the cat's ascent is unique in a way that contributes to the formation of this special gesture form. For the particular phenomena of gesture-speech association that we are attempting to account for here, we believe that an explanation that specifies significant semantic integration of the two modalities best fits the data. We do not provide a theory here of how grammatical or lexical structures of spoken Spanish achieve this congruence. We note only that our data on synchrony and gesture form point to an account that specifies a conceptual linkage between manner and ground in motion event expression.

## 9. Conclusions

Researchers often pose the question: who are gestures for, the speaker or the listener? With respect to the account we have offered for the manner-expressive gestures of Spanish speakers, we note that gestures are for linguists. This is not meant facetiously. A message emerges from a comparison of the results of the main and follow-up studies reported here. It is that, while for native Spanish speaker-listener pairs, the spoken language itself provides data sufficient for inferences concerning manner of motion, for *non*-natives (linguists) observation of speech-synchronous gesture may be an additional necessity. Without such observations, we underestimate the extent of manner conceptualization during acts of motion event description; also, we are hampered in understanding how manner is integrated with motion event conception as a whole.

We learned that manner gestures are equally frequent in the three languages, which argues against gesture-speech compensation. The alternative mentioned in the Introduction, that gesture and speech jointly highlight manner content, is not a claim that thinking-for-speaking with regard to manner is cross-linguistically invariable. We concur with Slobin (1998:2) that,

[...] descriptions of manner of—in various sorts of discourse—are to a large part determined by the *lexicalization patterns* of the language; and, as a consequence, thinking for speaking varies systematically on the basis of such patterns. (emphasis in original)

What we propose is that the systematicity of the variation across Spanish and languages like English and Chinese in manner conceptualization is not of the less-*versus*-more variety. Rather, it is variation with respect to the dynamic interplay of the various components of motion during formation of utterances that are both richly expressive and rhetorically appropriate.

Further examination of the expression of components of motion in gesture, along the lines of Slobin and colleagues' extensive and detailed analyses of speech in natural discourse contexts, is needed before the hypothesis offered here is solidly supported. We hope we have persuaded some concerning a methodological point; namely, the value of very close assessments of gesture-speech temporal and semantic synchrony. This is one of the important innovations of the

Kendon (1972, 1980) and McNeill (1992) research paradigms. We believe that such analyses have great potential for studies hoping to make theoretical claims regarding the role of visual and mental imagery in language production, the significance of gesture data in assessments of speaking-associated conceptualization, and the way languages structure meaning.

## References

- Duncan, Susan. 2002. Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture* 2(2): 183-206.
- Hadar, Uri and Brian Butterworth. 1992. Iconic gestures, imagery and word retrieval in speech. *Psychological Review* 99: 116-148.
- Kendon, Adam. 1972. Some relationships between body motion and speech: An analysis of an example. In A. Siegman and B. Pope (eds.) *Studies in Dyadic Communication*. New York: Pergamon Press.
- Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the process of utterance. In M.R. Key (ed.) *The Relationship Between Verbal and Nonverbal Communication*. The Hague: Mouton.
- Kendon, Adam. 2000. Gesture and speech: Unity or duality? In D. McNeill (ed.) *Language and Gesture*. Cambridge: Cambridge University Press.
- Kita, Sotaro. 2000. How representational gestures help speaking. In D. McNeill (ed.) *Language and Gesture*. Cambridge: Cambridge University Press.
- McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- McNeill, David. 1997. Imagery in motion event descriptions: Gestures as part of thinking-for-speaking in three languages. *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society.
- McNeill, David. 2000. Analogic/analytic representations and cross-linguistic differences in thinking for speaking. *Cognitive linguistics* 11(1/2): 43-60.
- Mueller, Cornelia. 1994. Semantic structure of motional gestures and lexicalization patterns in Spanish and German descriptions of motion-events. *Proceedings of the Thirtieth Annual Meeting of the Chicago Linguistic Society*, Vol. 1, 281-295.
- Nobe, Shuichi. 1996. Representational gestures, cognitive rhythms, and acoustic aspects of speech: A network/threshold model of gesture production. Ph.D. dissertation, University of Chicago.
- Slobin, Daniel. 1995. Learning to think for speaking: Native language, cognition, and rhetorical style. In J. Gumperz & S. Levinson (eds.) *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press.
- Slobin, Daniel. 1996. Two ways to travel: Verbs of motion in English and Spanish. In M. Shibatani and S. Thompson (eds.) *Grammatical Constructions: Their Form and Meaning*. Oxford: Clarendon Press.

Susan Duncan

- Slobin, Daniel. 1998. Verbalized events: A dynamic approach to linguistic relativity and determinism. *Working paper no. 449, LAUD symposium, Series A: General and theoretical*. Essen, Germany: Linguistic Agency University-GH. / 2000. In S. Niemeier and R. Dirven (eds.) *Evidence for Linguistic Relativity*. Amsterdam: John Benjamins.
- Slobin, Daniel and Nini Hoiting. 1994. Reference to movement in spoken and signed languages: Typological considerations. In S. Gahl, A. Dolbey, and C. Johnson (eds.) *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society.
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (ed.) *Language Typology and Syntactic Description. Vol. III. Grammatical Categories and the Lexicon*. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1991. Path to realization: A typology of event conflation. In L.A. Sutton, C. Johnson, and R. Shields (eds.) *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society.
- Tuite, Kevin. 1993. The production of gesture. *Semiotica* 93(1/2): 83-105.

Susan Duncan  
University of Chicago  
5848 S. University Ave, #210  
Chicago, IL 60637

# **The Development of Gesture, Speech, and Action as Communicative Strategies**

BARBARA F. KELLY

*University of California, Santa Barbara*

## **0. Introduction**

This paper examines relations between young children's gestures, utterances, and related actions, and the corresponding responses of adult caregivers. Previous research has shown that gesture and speech become integrated during the one-word period (Goldin-Meadow 1993). Prior to this integration, children make use of gestures or single words as a communicative strategy (Bates 1979, Butcher and Goldin-Meadow 2000). Integrating strategies allows a gesture to be used to direct a recipient's gaze toward a particular person, place, or thing, while words can be used to identify specific objects (Goldin-Meadow, Wein, and Chang 1992, Greenfield and Smith 1976). In this paper, I suggest that children employ *action* as a communicative strategy, together with speech and gesture. I show that when a child uses two, or three combined strategies (combinations of gesture, speech, action) they receive a higher percentage of responses from a recipient (i.e. acknowledgements of a communication) than when a single strategy is used.

## **1. Background**

Research in language acquisition revolves primarily around the question of how young children learn to both produce and comprehend spoken utterances. Prior to the onset of speech, much of the language development research focuses on the ways that preverbal behaviors are related to the development of speech. In language development, gestural communication has an important role from the beginning. Gesture is seen as a precursor to verbal language development (Dore 1975, Acredolo and Goodwyn 1988). For example, babbles of Japanese babies have been found to be more syllable-like and canonical when they wave their arms and open and close their hands (Masataka 2000), and early motor activities such as reaching and grasping have been linked to gestural babbling (Carter 1979, Pettito and Marentette 1991). Further activities, such as an infant's ability to follow another person's gaze (Churcher and Scaife 1982) or a pointing gesture (Butterworth and Grover 1990) have also been seen as crucial in the infant's path

of development toward more structured gestural actions and toward verbal development (Sachs 1993).

In single-word speech, young children systematically combine early utterances with non-verbal acts. These often express associations between two different elements about which a child wants to communicate with relation to a single event. As children's spoken language skills develop, they rely less heavily on the use of gesture in communication than they do on spoken language skills (Carpenter, Naigle, and Tomasello 1998). However, gesture has been found to lighten the cognitive load in interaction (Goldin-Meadow this volume), and the communication skills children use when they employ gesture as a means of communication are an important basis for their later verbal development, particularly beyond the one-word stage.

Several researchers have examined response types that may occur when a child produces a point (Jones 2000, Wootton 1997). First, there could be a non-response, where the caregiver does not acknowledge the child's communicative gesture. This may cause the child to attempt to reinitiate the sequence. A second possibility is that the caregiver might clarify what the point is doing. Finally, a caregiver may make a guess about what the point is doing. In this paper, I am interested in each of these categories at a somewhat broader level. I am interested in non-responses versus responses, regardless of whether those responses are clarification or attempted guess responses. This paper asks the following questions: What types of strategies typically elicit a caregiver's response as opposed to non-response? Do combinations of strategies elicit more responses than use of single strategies? If so, which combinations of strategies elicit the most response?

## **1.1. Method**

### **1.1.1. Subjects**

Data for this study are from an audiovisual corpus collected one hour weekly over an 18 month period in a day care center in Santa Barbara. The study focuses on four children taped from around age 12-16 months together with their caregivers.

### **1.1.2. Data coding**

The data were coded for several communicative strategies. A communicative strategy is the means by which a child initiates a communication. For each strategy coded, this typically involves the child looking at the caregiver and then moving their gaze toward a target object while using the strategy. The data were coded for three distinct communicative strategies. These are:

- *Vocalizations*, which may be words or proto-words (these are children's idiosyncratic but consistent attempts at producing adult words; for example, one child in the corpus uses [m★] to indicate that she wants a drink).
- *Gestures*, which are gestures directed at another individual. Two gesture types were coded according to coding laid out in Goldin-Meadow and Mylander (1984), including:

*Gesture, Speech, and Action as Communicative Strategies*

*Point*: a movement of the index finger or outstretched hand toward an object.

*Gimme*: holding out an open palm as if to receive an object; the palm may be turned upwards or sideward.

- *Actions*, which are movements toward or away from an object of attention in an effort to either grasp it or avoid it. Actions that are considered communicative are those where the child indicates an intention of an initiation to communicate. This is typically established through an engaged eye gaze to the caregiver. Not all actions are communicative; the additional cue of eye gaze must also be present for an action to be coded as communicative.

**2. Results and discussion**

Results indicate that across the four children single strategies were used most frequently, with 78 instances of a single strategy use, followed by 76 instances of a combination of two strategies, and 23 instances of use of three combined strategies. Data was coded by myself and a second coder. Cohen’s kappa was significant at  $p < .001$ ,  $k = .868$ , indicating high inter-coder reliability.

Table 1 below presents the results relating to the first question regarding which types of strategies typically elicit a caregiver’s response as opposed to non-response. The figures indicate that speech and gesture strategies elicit the most responses.

QU 1: WHAT TYPES OF STRATEGIES TYPICALLY ELICIT A CAREGIVER’S RESPONSE?

| <i>strategies</i>   | <i>speech</i> |     | <i>gesture</i> |     | <i>action</i> |     | <b>Total</b> |     |
|---------------------|---------------|-----|----------------|-----|---------------|-----|--------------|-----|
|                     | n             | %   | n              | %   | n             | %   | n            | %   |
| <b>response</b>     | 16            | 42  | 12             | 33  | 0             | 0   | 28           | 100 |
| <b>non-response</b> | 22            | 58  | 24             | 67  | 4             | 100 | 50           | 100 |
| <b>Total</b>        | 38            | 100 | 36             | 100 | 4             | 100 | 78           | 100 |

*Table 1.* Single communicative strategies across caregiver response

The highest percentage of caregiver responses to single strategies occurred when the child used a speech strategy. 42% of single utterances received a caregiver response. This was followed by 33% of single gestures receiving a response. No single actions garnered caregiver responses.

The small percentage of single action strategies and the lack of response to this strategy suggests that when it is the sole communicative strategy action does not have the same communicative value as gesture and speech. It is neither used by the child nor adhered to by the caregiver in the same way that the alternative strategies are. However, it must be noted that the use of action is not always an available option for children. Young children are often involved in activities that do not give them the freedom to move toward or away from an object they wish to

initiate a communication about. For example, one such incident involved a low-flying airplane near the day care center. Several children pointed and vocalized about this distant object, which they were not able to move toward. Another common instance of restricted movement is when a child is seated at a table and their chair is pulled close to the table edge, thus restricting their ability to move toward an object they are gesturing and or vocalizing about.

QU 2: DO COMBINATIONS OF STRATEGIES ELICIT MORE RESPONSES THAN USE OF SINGLE STRATEGIES?

Moving on to the second question of the paper, an investigation of whether combinations of strategies elicit a higher percentage of responses than use of single strategies, data indicates that the use of three combined strategies resulted in the highest percentage of caregiver responses, as seen in Table 2 below:

| <i># of strategies</i> | <i>one</i> |     | <i>two</i> |     | <i>three</i> |     |
|------------------------|------------|-----|------------|-----|--------------|-----|
|                        | n          | %   | n          | %   | n            | %   |
| <b>response</b>        | 28         | 38  | 60         | 79  | 22           | 96  |
| <b>non-response</b>    | 50         | 62  | 16         | 21  | 1            | 4   |
| <b>Total</b>           | 78         | 100 | 76         | 100 | 23           | 100 |

Table 2. Caregiver responses across combinations of categories

Table 2 indicates that 96% of three combined strategies were responded to by the caregiver. The use of two combined strategies resulted in a response 79% of the time, while use of a single strategy was responded to 38% of the time. The figures indicate a higher response rate for combinations of strategies than for single strategies. The frequency of response is highest for the combination of all three strategies, indicating that combinations of strategies elicit more responses than use of single strategies.

The use of different communicative strategies by the children is illustrated in the transcript in (1) below. This transcript shows single strategies and combinations of strategies in an interaction between a caregiver and two children. In the scene presented here, Chera (aged 1;3) is sitting eating lunch and the caregiver is sitting near her. Beside the table Chera is eating at, there is a smaller table with her lunch containers and bottle. The transcript begins when Brailey (aged 1;0) moves toward the table and vocalizes. The number of strategies in this short interaction is not indicative of the frequency of strategies as a whole across the database. Meal times tend to be particularly fruitful places for finding both pointing and gimme gestures. This is largely due to the fact that for many children other children’s food is usually more interesting and sought after than their own.

In the transcript below, the numbers in the left hand column represent strategy sequence numbers and relate to the adjacent strategy labels in the column beside them. These in turn relate to the information in the columns headed Brailey and

### *Gesture, Speech, and Action as Communicative Strategies*

Caregiver. The numbers in the # column refer to the total number of strategies for each sequence. I have placed the category of *gaze* in the 'strategy' column, although, as mentioned in 2.2, *gaze* is not a strategy. A *gaze* is used by the child to mark the intention of an initiation to communicate, which is the first step in coding each of the communicative strategies.

In sequence 1 in the following transcript, Brailey uses a single communicative strategy; an utterance "de" and receives no response from the caregiver. In 2, he uses two strategies, an utterance "yeya" which combines with a gesture (a point toward food containers on the table). Here he receives the response "That's Chera's lunch". In 3, Brailey points to the container again and combines this with an utterance "da" and the caregiver responds with "That's Chera's". In 4, Brailey then moves his focus to the smaller table the containers are on, and points to that, combining this gesture with the utterance "da". This receives a minimal response of "Ahem" from the caregiver, who then turns to focus on Chera (who, incidentally, is pointing or reaching toward her food throughout this interaction, but is not responded to until she verbalizes). Brailey then moves away from the table and is out of camera range for 11 seconds. He moves back into range and as can be seen in 5 he points to and touches a food container while saying "da". This combined strategy receives no response from the caregiver; however, in 6, when Brailey adds the action of attempting to pick up the container, the caregiver shifts her attention from Chera, who still has her hand out in a reaching gesture, and turns around to face Brailey. She then responds by physically intervening between him and the containers while saying "That's Chera's, I don't want you to pick it up". In 7, when Brailey continues to combine three communicative strategies, the caregiver again physically intervenes while asserting the fact that "That's Chera's too". In 8, Brailey returns to combining just two strategies, utterance and action, because his action involves drawing the container to him with both hands. The object that has been the referent of his gestures all along is now in his grasp and he therefore has no need for the pointing gesture he has been using. This combined action-plus-utterance strategy leads to the fullest spoken response thus far when the caregiver says "Brailey, I don't want you to play in Chera's food". She then takes the final action of the sequence, which is to move Brailey away from the table where the lunch containers are.

(1)

| Time                    | Strategy                                      | Brailey (12 months)                                                                                        | Caregiver                                                                                                     | # |
|-------------------------|-----------------------------------------------|------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|---|
| 35:02                   | 1: gaze<br>utterance                          | on lunch containers<br>de                                                                                  | <i>No response</i>                                                                                            | 1 |
| 35:05<br>35:05          | 2: gaze<br>utterance<br>gesture               | on one lunch container<br>yeya [Chera]<br>points to container                                              | That's Chera's lunch<br>(focuses on other child)                                                              | 2 |
| 35:08<br>35:10          | 3: gaze<br>utterance<br>gesture               | still on lunch container<br>de<br>points to container<br>(finger touches it)                               | That's Chera's                                                                                                | 2 |
| 35:13<br>35:13          | 4: gaze<br>utterance<br>gesture               | moves to table<br>da<br>points to table<br><br>[moves out of view]                                         | uh-hm<br>(focus moves to other<br>child)                                                                      | 2 |
| 35:31<br>35:31          | 5: gaze<br>utterance<br>gesture               | on food containers<br>da<br>points to and touches<br>container                                             | <i>No response</i><br>(focus is on other child)                                                               | 2 |
| 35:33<br>35:34<br>35:35 | 6: gaze<br>gesture<br><br>utterance<br>action | on one container<br>pointing<br><br>da<br>touching then picking up                                         | <i>physical intervention</i><br>That's Chera's; I don't<br>want you to pick it up.                            | 3 |
| 35:42<br>35:42<br>35:42 | 7: gaze<br>utterance<br>gesture<br>action     | on different container<br>da .. da ... da<br>points to new container<br>attempting to pick up<br>container | <i>physical intervention</i><br>That's Chera's too<br>mhm                                                     | 3 |
| 35:46<br>35:47          | 8: gaze<br>action<br><br>utterance            | on food container<br>draws container to him<br>with both hands<br><br>da                                   | <i>physical intervention</i><br>(moves child away)<br><br>Brailey I don't want you<br>to play in Chera's food | 2 |

*Gesture, Speech, and Action as Communicative Strategies*

These interaction sequences nicely illustrate children’s use of single and multiple strategies and the types of responses these strategies receive from caregivers. They show how the use of different strategies and combinations of strategies can result in different caregiver responses. The interaction between Brailey and the caregiver throughout the time that Chera is gesturing toward her food indicates a circumstance where a child using just one single strategy forfeits the attention of the caregiver, whose focus moves to the child who is using combined strategies.

It is possible to see this interaction as an escalation of strategies due to Brailey’s persistence in trying to get to the containers, and this persistence as being the motivation for the caregiver’s response, not the use of multiple strategies. Table 3 below presents a partialling out of interaction strategies on the basis of whether the caregiver is attempting to monitor or moderate a child’s behavior. Restricted strategies are those where the caregiver responds to the child with a direct comment on their behavior, as in (1) above when the caregiver says “Brailey I don’t want you to play in Chera’s food”.

| <i># of strategies</i> | <i>one</i> |     | <i>two</i> |     | <i>three</i> |     | <b>Total</b> |     |
|------------------------|------------|-----|------------|-----|--------------|-----|--------------|-----|
|                        | n          | %   | n          | %   | n            | %   | n            | %   |
| <b>restricted</b>      | 1          | 25  | 2          | 7   | 1            | 5   | 4            | 8   |
| <b>non-restricted</b>  | 3          | 75  | 26         | 81  | 22           | 95  | 51           | 92  |
| <b>Total</b>           | 4          | 100 | 28         | 100 | 23           | 100 | 55           | 100 |

*Table 3. Restricted versus non-restricted action strategies*

The distinction highlighted in Table 3 is between those interactions when a caregiver is explicitly attempting to monitor a child’s actions in some way and those when the child’s actions are not an issue. Results in Table 3 indicate that of the 55 action strategies, 4 (7%) were coded as being responded to with a behavior monitoring comment. The transcript in (2) below illustrates use of multiple strategies and caregiver response when the response is not for behavior monitoring purposes.

In the following transcript, Chera (aged 1;3) is sitting eating lunch with some other children and the caregiver is sitting across the table from her. In sequence 1, Chera sees some snacks on the table and points toward them while looking at the caregiver who is facing her, but gives no response. In 2, she then sees a bag of pretzels near the caregiver and reaches toward it. She receives no response from the caregiver and moves her attention to a container of pasta. In sequence 3, she points to the container saying “baba”. Chera has escalated her use of strategies in this sequence, using both gaze and gesture, and she receives a response from the caregiver, who encourages her to try some pasta. In 4, Chera then holds her gaze on the pasta and there is a 7 second pause between the caregiver’s suggestion to try some and Chera’s response with “ba”, which she says as she attempts to stand up and simultaneously reach toward the pasta.

(2)

| Time  | Strategy                           | Chera (15 months)                         | Caregiver                                                           | # |
|-------|------------------------------------|-------------------------------------------|---------------------------------------------------------------------|---|
| 06:58 | 1: <b>gaze</b><br><b>gesture</b>   | on caregiver<br>points to snacks on table | <i>No response</i>                                                  | 1 |
| 07:02 | 2: <b>gaze</b><br><b>gesture</b>   | on bag of pretzels<br>reaches for bag     | <i>No response</i>                                                  | 1 |
| 07:06 | 3: <b>gaze</b><br><b>utterance</b> | moves to pasta container<br>baba          | Do you know what this is? This is pasta. Do you want to try a bite? | 2 |
| 07:06 | <b>gesture</b>                     | points to container of pasta              |                                                                     |   |
| 07:13 | 4: <b>gaze</b><br><b>utterance</b> | on pasta in front of her<br>ba            | No?                                                                 | 3 |
| 07:13 | <b>gesture</b>                     | points to pasta                           |                                                                     |   |
| 07:14 | <b>action</b>                      | attempts to stand up and reach for pasta  |                                                                     |   |

In this interaction we can see an increase in the number of strategies used by the child, although the caregiver responses are very different from those illustrated in (1) above. The strategy used in this transcript reflects the findings presented in Table 2 above with relation to caregiver responses across communicative strategies. The number of strategies used by Chera increased with each sequence; however, there was no corresponding behavior-monitoring by the caregiver, such as the monitoring of Brailey’s behavior in (1) above. Caregiver response is clearly not just triggered by persistent behavior such as Brailey’s in (1), as, when we look at (2), we can see that Chera’s multi-strategy use did not trigger a behavior-based comment and yet she still received responses when she used more than one strategy. The caregiver responded to the children’s use of communicative strategies regardless of their behavior.

QU 3: WHICH COMBINATIONS OF STRATEGIES ELICIT THE MOST RESPONSE?

A tally of the most frequent strategy combinations in the transcripts in (1) and (2) suggests that the most frequent combination of strategies is utterance and gesture. This is not, however, reflective of the overall results of the study, as we can see when we look at results for the third question regarding which combinations of strategies gain the most response. Table 4 below indicates that the combination most frequently responded to was speech and action, with 87% of speech and action strategies receiving a caregiver response, followed by action and gesture with 83% and speech and gesture with 75%.

*Gesture, Speech, and Action as Communicative Strategies*

| <i>strategies</i>   | <i>speech/gesture</i> |     | <i>speech/action</i> |     | <i>gesture/action</i> |     |
|---------------------|-----------------------|-----|----------------------|-----|-----------------------|-----|
|                     | n                     | %   | n                    | %   | n                     | %   |
| <b>response</b>     | 36                    | 75  | 14                   | 87  | 10                    | 83  |
| <b>non-response</b> | 12                    | 25  | 2                    | 13  | 2                     | 17  |
| <b>Total</b>        | 48                    | 100 | 16                   | 100 | 12                    | 100 |

*Table 4. Combined communicative strategies across caregiver response*

The results in Table 4 indicate that there were minimal differences in response frequency between the combined strategy pairs, and these are not statistically significant ( $\chi^2$  (df=2, n=76) = 3.78  $p < .05$ ).<sup>1</sup> Regardless, there are several ways of combining these pairs, all of which are more effective than single strategy uses.

This finding indicates the importance of children’s use of action alongside speech and gesture as a means of eliciting recipient responses to communicative strategies. The two primary combined strategies included action, which we noted with relation to question one, is not an important communicative strategy on its own. It is, however, important as a strategy in combination with either speech or gesture and more so with them both. Although action is not always an available choice for young children, when it is used, it is a valuable resource in initiating communication.

### **3. Conclusion**

Overall, results from this study indicate that the use of an integrated system results in a higher percentage of responses from a caregiver than the use of single communicative strategies. Children received a higher percentage of responses from a recipient when they used two or three combined communicative strategies than when they used a single strategy. In their use of two combined strategies, children received the highest response using a speech and action strategy; however, response differences across strategy pairs were not significant.

Results presented in this study indicate that caregiver response differs across different types and frequencies of communicative strategies. This finding provides insight into a motivating factor of children’s development from using independently functioning gesture and speech systems to use of an integrated gesture/speech/action modality.

### **References**

Acredolo, L. and S. Goodwyn. 1988. Symbolic gesturing in normal infants. *Child Development* 58: 450-466.  
 Bates, E. 1979. *Language and context*. New York: Academic Press.

<sup>1</sup> Expected frequency was < 5 in two cells.

- Butcher, C. and S. Goldin-Meadow. 2000. Gesture and transition from one- to two word speech: When hand and mouth come together. In D. McNeill, ed., *Language and gesture*, 235-58. Cambridge: Cambridge University Press.
- Butterworth, G. and L. Grover. 1990. Joint Visual Attention, Manual Pointing, and Preverbal Communication in Human Infancy. In M. Jeannerod, ed., *Attention and Performance XIII*. Hillsdale, NJ: Lawrence Erlbaum and Associates.
- Carpenter, M., K. Nagel, and M. Tomasello. (1998). Social cognition, joint attention and communicative competence from 9-15 months of age. In *Monographs of the society for research on child development*, 255. 63(4).
- Carter, A. 1978. The development of systematic vocalizations prior to words: A case study. In N. Waterson and C. E. Snow, eds., *The development of communication*, 127-138. Chichester: Wiley.
- Churcher, J. and M. Scaife. 1982. How infants see the point. In G. E. Butterworth and P. Light, eds., *Social Cognition*, 110-136. Hassocks. Harvester Press.
- Dore, J. 1975. Holophrases, speech acts, and language universals. *Journal of Child Language*, 2: 20-40.
- Goldin-Meadow, S. 1993. When does gesture become language? A study of gesture used as a primary communication system by deaf children of hearing parents. In K. R. Gibson and T. Ingold, eds., *Tools, language and cognition in human evolution*, 63-85. Cambridge: Cambridge University Press.
- Goldin-Meadow, S. 2000. Beyond words: The importance of gesture to researchers and learners. *Child Development Special Issue: New Direction for Child Development in the Twenty-First Century*. 71: 231-139.
- Goldin-Meadow, S. and C. Mylander. 1984. Gestural communication in deaf children: The effects and non-effects of parental input on early language development. *Monographs of the Society for Research in Child Development*. 49: 1-121.
- Goldin-Meadow, S., D. Wein, and C. Chang. 1992. Assessing knowledge through gesture: using children's hands to read their minds. *Cognition and Instruction* 9: 201-219.
- Greenfield, P. and J. Smith. 1976. *The structure of communication in early development*. New York: Academic Press.
- Jones, S. 2000. Pointing as a recognizable communicative action in toddler caregiver interaction. M.A thesis, University of California, Santa Barbara.
- Masataka, N. 2000. The role of modality and input in the early stages of language acquisition: Studies of Japanese sign language. In C. Chamberlain, J. Morford, and R. Newberry, eds., *Language acquisition by eye*. New Jersey: Lawrence Erlbaum and Associates.
- Petitto, L. A. and P. A. Marinette. (1991). Babbling in the manual mode. *Science* 251:1483-1496.
- Sachs, J. 1993. The emergence of intentional communication. In J. Berko-Gleason, ed., *The development of language*. New York: Macmillan.
- Wootton, A. 1997. Interaction and development of the mind. In P. Drew, M. H. Goodwin, J. Gumperz and D. Schiffrin, eds., *Studies in Interactional Sociolinguistics 14*. Cambridge: Cambridge University Press.

# **An Analysis of Shifts in Participation Roles in Japanese Storytelling in Terms of Prosody, Gaze, and Body Movements\***

CHISATO KOIKE

*University of California, Los Angeles*

## **0. Introduction**

When we tell a story to others about our experiences, we all become actors and actresses. We not only present the story line as a narrator but also we act out what people said or did to invoke the scenes of the past and vividly reconstruct them in front of our audience. In other words, we have to shift our roles as narrator and characters as we tell the story. In this paper, I will analyze shifts in participation roles in a storytelling segment of a casual conversation among Japanese friends, focusing on direct quotations, or more precisely “demonstrations” (Clark and Gerrig 1990), in terms of prosody, gaze, and body movements, as well as syntactic structure.

In his analysis of participation framework, Goffman (1981) proposed the notion of “footing,” by which a speaker changes his/her roles of utterance production as “animator,” “author,” or “principal.” Participation framework has also been analyzed from more interactive perspectives. Goodwin (1984) pointed out that in the course of storytelling the participants interactively organized their talk, as well as visibly displayed actions such as gaze and body movements. Goodwin (1997) found that not only a narrator but also her recipients displayed evaluation vocally as well as non-vocally and coordinated the evolving storytelling through “byplay.” Building on these studies of participation framework, I will examine how the storyteller strategically assumes several different roles using prosody, gaze, and body movements and how she shifts not only her roles but also roles of other participants in the conversation in order to involve them in the story of her own experience.

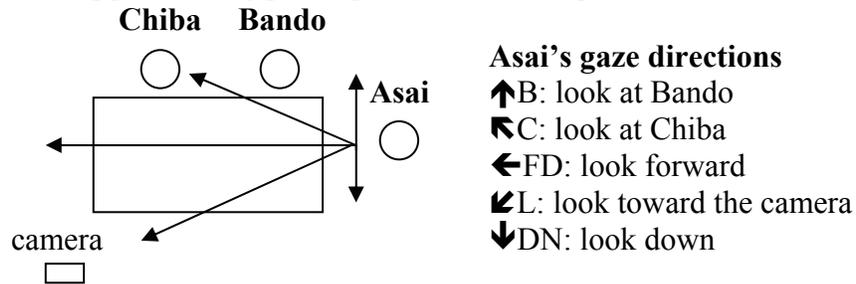
---

\*This is a revised version of my qualifying paper. I am most grateful to Professor Charles Goodwin for his valuable advice and suggestions in preparing for this paper. I also would like to express my gratitude to Professor Noriko Akatsuka and Professor Shoichi Iwasaki for their support. My great appreciation is extended to Matt Burdelski and Tess Post for correcting my English in this paper and to my audience at the BLS 27<sup>th</sup> Annual Meeting for giving me inspiring comments and suggestions.

**1. Data and Methods**

The conversational data examined in this study are contained in an approximately one-minute storytelling segment of a Japanese casual face-to-face video-taped conversation among three intimate female friends in their early twenties—Asai, Bando and Chiba, who were studying ESL at a university in the United States at the time of recording. I videotaped their conversation in a small university classroom. The figure in (1) shows the seating positions of the participants and Asai’s gaze directions. I put Asai’s gaze direction and description of her body movements under the gloss of Asai’s utterances in transcription.

(1) *Seating positions of participants and Asai’s gaze directions*

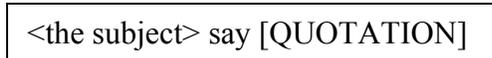


**2. Analysis**

**2.1. Quotation in Japanese conversation**

First, I would like to explain briefly about the Japanese quotation pattern compared that of English. In English, the subject and the verb *say* come before the quotation, as illustrated in (2).

(2) *English quotation pattern*



In (3), Ann is quoting what her husband Don said to her, while she is telling a story to her dinner hosts about a certain kind of wallpaper in her friend’s new house. Ann prefaces her direct quotation of Don’s utterance with *Do(h)n said* (the subject and the verb *say*).

(3) *Quotation in English* (Goodwin 1984:226)

13 Ann: Do(h)n said, (0.3) dih- did they ma:ke you take this wa(h)llpa(h)er? er(h) didju pi(h)ck i(h)t ou(h)t.

In Japanese, the quotative particles *to*, *tte* or *toka*<sup>1</sup> are used to mark quoted speech or thoughts. These quotative particles and *iu* ‘say’ come after the quota-

<sup>1</sup> *Tte* is an informal version of the quotative particle *to* and is used quite frequently in colloquial language. *Toka* ‘or something’, originally a combination of a quotative particle *to* and a question

tion, as illustrated in (4). Therefore, a listener cannot project the beginning of quotation in Japanese.

(4) *Japanese quotation pattern*<sup>2</sup>

|               |           |             |             |           |
|---------------|-----------|-------------|-------------|-----------|
| <the subject> | <b>ga</b> | [QUOTATION] | <b>to</b>   | <b>iu</b> |
|               | subject   |             | <b>tte</b>  | say       |
|               | particle  |             | <b>toka</b> |           |
|               |           |             | quotative   |           |
|               |           |             | particle    |           |

In (5), Eko is quoting what Jane said to her while she is telling a story about Jane's popularity. Eko prefaces her quotation with *Jane ga* 'Jane' (the subject + the subject particle) and marks the end of the quotation by *toka itte* 'or something say and' (the quotative particle + *itte* 'say-and'<sup>3</sup>).

(5) *Japanese quotation* (from my data corpus)

Eko: (0.4) *Jane ga, ima Daniel:: ga:, kuruma de:,*  
 Jane SUB now Daniel SUB car by  
 (0.4) *mukae ni kite-kureta toka itte,*  
 welcome to come-AUX QT say-and  
 'Jane says now Daniel came to pick me up by car or something, and'

A speaker also sometimes omits *to iu* (the quotative particle + 'say') after the quotation, so that the end of quotation is also sometimes not clear in terms of syntactic structure. In addition, since in Japanese conversation a speaker often does not indicate the subject of a quotation, syntactic structure does not clearly indicate who said the quoted speech as in (6).

(6) *Simplified Version of (8)*

89 Asai: [U:n. soide ittara:,  
 Yeah. And then when (I) went (to the window),  
 → *kon↑ni↓chiwa:: hoomusu↑te:i:?* [((laughter))  
*konnichiwa* 'hello'. *hoomusutei?* 'home-stay?'  
 90 Chiba: [((laughter))  
 91 Bando: [((laughter))=  
 92 Chiba: =*Saki ni kika[rete shi(h)ma(h)tta.*  
 (You) ended up being asked first.

particle *ka*, is used as a quotative particle to indicate "the speaker's uncertainty about the quoted report" (Makino and Tsutsui 1986:489).

<sup>2</sup> [The subject + *ga*] can be placed after [quotation + *to*].

<sup>3</sup> *Itte* 'say-and' is the gerund form of the verb *iu* 'say'. The fundamental function of the gerund form of verbs, adjectives and copula is to link sentences, clauses, and phrases.

- 93 Asai: [U(h)n. hh  
Yeah.  
94 Chiba: ((laughter))  
→ 95 Asai: (0.6) *Hai*. [t(h)ok(h)a i(h)tte:,  
*Hai* ‘Yes’ or something, (I) said and,

In (6), Asai quotes the immigration official’s speech in 89A and Asai’s own speech in 95A, but she does not indicate the subject of quoted speech for either utterance. She also omits *to iu* (the quotative particle + ‘say’) after the quoted speech of the immigration official in 89A *konnichiwa. hoomusutei?* ‘hello. home-stay?’<sup>4</sup> This raises the question of how a speaker indicates different voices to depict a story in which there is more than one character without explicitly stating the subject of the quotations.

## 2.2. Multiple devices for multiple voices: prosody, gaze and body movements

In this section I will demonstrate how the storyteller manipulates her voice, gaze, and body movements to shift among four different roles in her story. Prior to Asai’s storytelling, the participants were talking about the questions they were asked at an international airport immigration window when they entered the United States. For example, when foreigners land at Los Angeles International Airport, an immigration official checks their passports and entry visas and usually asks them several questions in English about purpose and length of stay. In the transcribed excerpt, Asai is telling a story about her experience at an immigration window when she first came to the United States. She is telling her co-participants that she practiced the answers to the questions she thought she would be asked in English by the immigration official. However, since the immigration official asked her questions in Japanese with an American English accent, Asai ended up answering those questions in Japanese.

During her storytelling, she plays four different roles: narrator, Asai herself, and two immigration officials. As I mentioned above, in Japanese conversation, syntactic structure not only does not always indicate who the subject of the quotation is but also where the beginning and the end of the quotation are. However, the storyteller in my data strategically used prosody, gaze, and body movements to mark the beginning of a quotation and to indicate a shift in her roles as narrator, the immigration officials, and Asai herself as in (7) and (8).

Prior to Excerpt (7), Asai told her co-participants that there were about three long lines of people who were waiting at the airport immigration windows and a fourth line was added while she was waiting. In (7), she is saying that she was told to line up in the new fourth line by the immigration official outside the desk,

---

<sup>4</sup> *Hoomusutei* ‘home-stay’ is a Japanese coined word from English, which literally means “staying in a person’s home,” but it usually refers to “staying with an American family for a certain period to study English.”

*An Analysis of Shifts in Participation Roles*

who was directing people to different lines. She indicates the role shift from the narrator to the immigration official by the use of pronouns and body movements.

(7) *The immigration official outside the desk directing people*

82 Asai: *Shitara*:

And then

↓DN

→

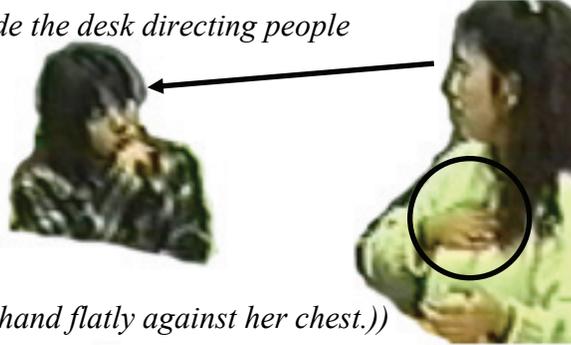
*atashi kara*:

I from

↑B

((Asai puts her right hand flatly against her chest.))

And then, from me



→

*omae kara*

you from

↑B

((Asai motions to lightly grab Bando's left upper arm with her right hand.))

from you



*kocchi narabe t(h)te i(h)wa(h)re(h)te*;

here line QT say-PASS-and

↑B

↖C

((Releasing Bando's arm, Asai motions as if she is bringing Bando's arm toward her and then away from her making a 90-degree curving motion.))

line up here, (I) was told, and,

In line 82A, she says *atashi kara*: 'from me' as she puts her right hand on her chest. Next, she says *omae kara* 'from you' as she lightly grabs Bando's left upper arm with her right hand, and then says *kocchi narabe t(h)te i(h)wa(h)re(h)te*: 'line up here, (I) was told, and' as Asai motions as if she is bringing Bando's arm toward her and then away from her making a 90-degree curving motion.

There is no significant change in prosody in line 82A except that she puts more emphasis on *atashi kara*: 'from me'. Her change from *atashi kara* 'from me' to *omae kara* 'from you' is not replacement as self-initiated self-repair in the same-turn. Rather, Asai is indicating the shift in roles from the narrator to the immigration official and clearly marking the beginning of the quotation by the use of pronoun *omae* 'you' and by body movements. In addition, not only is she quoting the immigration official's speech but she is also demonstrating what the

immigration official did to her by supplementing the information with the immigration official's body movements, i.e. grabbing Asai's arm and taking her to the new fourth line. I also contend that Asai's quotation here is not an exact direct quotation in a strict sense, because she translated what she thought the official said to her in English into Japanese.<sup>5</sup> It is worth noting that in 82A she did not mimic the immigration official's voice, as she did in lines 89A, 97A and 99A of Excerpt (8), where she exactly quotes what the immigration official said in Japanese.

Prior to Excerpt (8), she told to her co-participants that she was instructed to line up in the new fourth line and she kept practicing the expected answers in English feeling very nervous. In (8), Asai is telling what another immigration official at the desk asked her in Japanese and how she responded to him in Japanese at the immigration window. She indicates the beginning of the quotation and shifts in roles through prosody and gaze.

(8) *Asai and the immigration official at the desk*

89 Asai:[U:n. soide ittara:,  
 yeah and then went-if  
 ↑B ↖C  
 Yeah. And then when (I) went (to the window)

→ kon↑ni↓chiwa:: hoomusu↑te:i:? [((laughter))  
 hello home-stay  
 ↑B ↖C  
 ((Asai points at Bando with the index finger of  
 her right hand.))  
 konnichiwa. hoomusutei? 'hello. home-stay?' (In Japanese with an  
 American English accent)



<sup>5</sup> It is also interesting to note that Asai translated the English word *you* not into the more neutral word *anata* 'you' but into *omae* 'you', which is deprecatory except when a husband uses it to speak to his wife. Her word choice here reflects and indicates how she interpreted the immigration official's behavior and manner of talk.

*An Analysis of Shifts in Participation Roles*

90 Chiba: [((laughter))  
 91 Bando: [((laughter))=

92 Chiba: =*Saki ni kika[rete shi(h)ma(h)tta.*  
 ahead in ask-PASS ended-up  
 (You) ended up being asked first.

93 Asai: [U(h)n. hh  
 yeah  
 ↖C ↑B  
 Yeah.

94 Chiba: ((laughter))

→ 95 Asai: (0.6) Hai.  
 yes  
 ←FD ←  
 Hai. 'Yes'.



[t(h)ok(h)a i(h)tte.;  
 QT say-and  
 ↑B  
 or something, (I) said and,

96 Chiba: [((laughter))

→ 97 Asai: (0.6) Is-shuu↑kan?  
 one week  
 ↑B  
 ((Asai extends her right index finger toward Bando.☝))  
 Is-shuukan? 'One week?' (In JPN with an A-English accent)

*Ni-shuu*↑kan?  
 two weeks  
 ↑B  
 ((Asai extends her right index and middle fingers toward  
 Bando.☝))  
*Ni-shuukan?* 'Two weeks?' (In JPN with an A-English accent)

98 Chiba: *Ni[hongo de?*  
 Japanese in  
 In Japanese?

- 99 Asai: [San-shuu↑kan?  
 three weeks  
 ↑B ↖C ↑B  
 ((Asai extends her right index, middle and ring fingers toward Bando.))  
 San-shuukan? ‘Three weeks?’ (In JPN with an A-E accent)



>toka iu kara:,< un.  
 QT say so yeah  
 ↖C  
 or something, (he) says, so yeah.

(0.8) yon-shuuka(h)n. [((laughter))  
 four weeks



((in a softer voice))  
 ((Asai tilts her head slightly to the left, and then raises her right hand with four fingers extended to the side of her face.))  
 Yon-shuukan. ‘Four-weeks’.

- 100 Chiba: [((laughter))  
 101 Bando: [((laughter))

In 89A, first Asai is describing what she did as a narrator *soide ittara* ‘and then when (I) went (to the window)’. Then she says *kon↑ni↓chiwa::*. *hoomusu↑te:i:?* ‘hello. home-stay?’ as she turns her gaze back to Bando. She indicates her shift in roles from the narrator to the immigration official by mimicking his Japanese with an American English accent as well as by shifting her gaze towards Bando. She is also demonstrating the official’s body movements, i.e. his pointing gesture, when she says *hoomusutei* ‘home-stay’. In 95A, Asai is demonstrating how she responded to the official’s question in Japanese. It is

important to note that she uses different registers of *yes* in 93A and in 95A. When she agrees with Chiba and responds affirmatively in 93A, she uses a casual form *Un* ‘Yeah’ and when she quotes her response to the immigration official’s question in 95A, she uses a formal form *Hai* ‘Yes’.

In addition, in 95A, she clearly marks the boundary of her quotation by gaze shift. After responding to Chiba in 93A *U(h)n* ‘Yeah’, Asai shifts her gaze from Bando and looks forward in 95A. In 95A, after a 0.6 second pause, she quotes her own speech *Hai* ‘Yes’ while still looking forward. Then she says *to ka itte* ‘or something (I) said, and’ as she looks at Bando. In other words, her quotation, or more precisely, her demonstration starts from the beginning of the 0.6 second pause when she looks forward and ends just before she says *to ka itte* ‘or something (I) said, and’ when she looks at Bando. Asai’s demonstration of her response to the immigration official is not only marked by gaze shift but also laughter and facial expressions. In 93A, Asai responds to Chiba as she is laughing and smiling at Chiba and Bando. When Asai is demonstrating her response to the official, namely the 0.6 second pause and *Hai* ‘Yes’ in 95A, she looks serious without smile. Then when she exits her demonstration, she smiles at Bando as she says *toka itte* ‘or something (I) said and’ laughing.

In 97A and 99A, she also uses prosody, gaze shift, and body movements to indicate her shift in roles from the narrator to the immigration official and from the narrator to Asai herself respectively. In 97A and the first line of 99A, Asai looks at Bando, uses pointing gestures, and uses an American English accent to demonstrate how the immigration official asked Asai questions in Japanese with an accent. In the third line of 99A, Asai tilts her head slightly to the left and looks forward. Then, 0.8 seconds later she quotes her own speech *yon-shuukan* ‘four weeks’ as she demonstrates how she said it in Japanese using iconic pointing gestures, a softer voice, and a puzzled facial expression to indicate her confusion and perplexity. It should be noted that in 97A and in the third line of 99A, Asai uses the same strategy as in 95A. She clearly indicates shifts in her roles by gaze and starts her demonstration before her speech as shown by her gaze shift at the beginning of a 0.6 second pause in 97A and a 0.8 second pause in 99A.

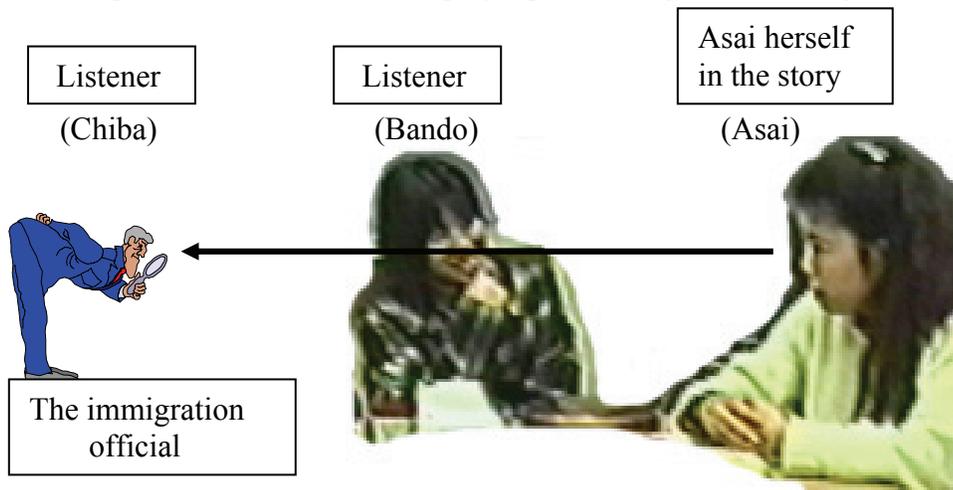
### **2.3. Shifting participation roles in storytelling**

Now analysis of Asai’s demonstration brings us to another question. In 89A, 97A, and the first line of 99A, when she is demonstrating the immigration official’s speech, why does Asai look at and point to Bando? Why does she look forward in 95A and look to the left in the third line of 99A when she is demonstrating her own speech? Isn’t her mimicking voice of Japanese with an American English accent enough to demonstrate the official’s speech? To investigate the storyteller’s gaze shift and pointing gestures in more detail, let us examine 82A, 89A, 97A, and 99A where Asai is demonstrating the immigration official’s speech.

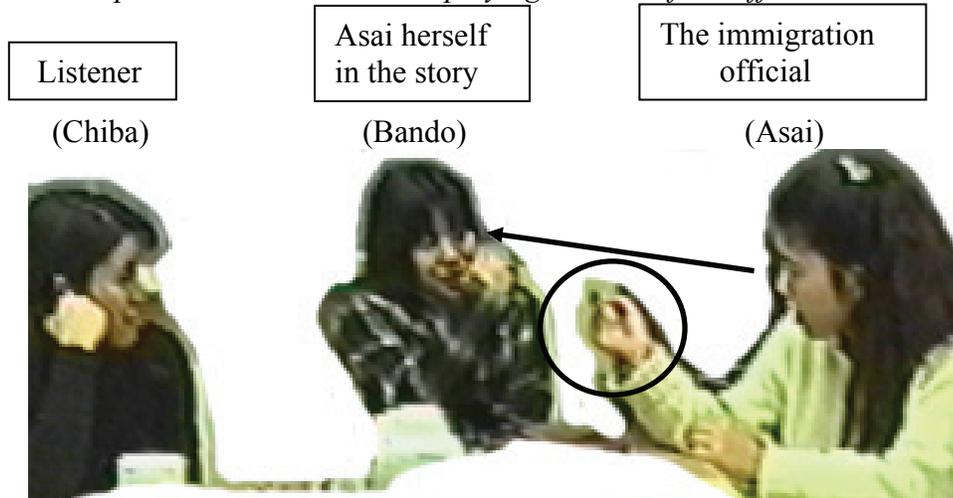
In 82A, as I mentioned earlier, Asai demonstrates what the immigration official said and did. Looking at Bando, Asai says *omae kara kocchi narabe* ‘from you line up here’ as she lightly grabs Bando’s left upper arm with her right hand,

and brings her right hand toward her and then away from her. Now who is the immigration official looking at and holding with his hand? It is Asai that he is looking at and holding. In other words, Asai as the immigration official is treating Bando as Asai herself in the story. In 89A, after describing the scene as a narrator, Asai looks at Bando again and demonstrates the immigration official's speech *konnichiwa. hoomusutei?* 'hello. home-stay?' in Japanese with an American English accent as she points at Bando with the index finger of her right hand. Here again, Asai treats Bando as Asai herself and visually creates the scene where the immigration official is talking with Asai.

(9) *Participation roles when Asai is playing the role of "Asai herself"*



(10) *Participation roles when Asai is playing the role of the official*



In 95A Asai looks forward to play the role of Asai herself and demonstrates how she says *Hai* 'Yes' in Japanese to the official in a serious and nervous look. Then in 97A and the first line of 99A, Asai goes back to the role of the official

and treats Bando as Asai herself again. Asai as the official, looks at Bando as Asai herself, and asks questions, *Is-shuukan? Ni-shuukan? San-shuukan?* ‘One week? Two weeks? Three weeks?’ in Japanese with an American English accent as she is using iconic pointing gestures toward Bando. In the third line of 99A, Asai plays the role of herself again. She looks forward and tilts her head slightly to the left and demonstrates her response in Japanese toward the immigration official.

The figures in (9) and (10) illustrate how Asai changes the participation roles in her storytelling. From the story preface to the climax of her story (99A), when she demonstrates her own speech, Asai constantly looks forward imagining that the immigration official is in front of her. She looks and points at Bando with her fingers when she demonstrates the immigration official’s speech as if she is the immigration official and Bando is Asai herself. Asai shifts, not only her roles as narrator, the immigration officials and Asai herself back and forth, but also Bando’s roles from a listener to Asai herself, back and forth. She uses Bando, who is proximate to Asai, as a story material, i.e. one of the characters in the story, in order to make the story lively, to depict the scenes more vividly and to involve co-participants in her story.

### **3. Conclusion**

In this paper, I analyzed how the storyteller used syntactic features, pauses, prosody, gaze, and body movements to quote, or more precisely demonstrate her own speech and others’ speech in a Japanese conversation among three female friends. I pointed out that in Japanese conversation it is difficult to project the boundary of a quotation in terms of syntactic features because *to iu* ‘the quotative particle + say’ comes after the quotation if it is stated at all. I also pointed out that in Japanese conversation a speaker does not often state the subject of quotations; therefore, syntactic structure does not indicate who spoke the quoted phrase.

However, as demonstrated in this study, the storyteller in my data strategically used multiple devices such as prosody, gaze, and body movements to shift her roles from narrator to the other three characters in her story back and forth. She mimicked voices of the characters. She looked forward when she played the role of herself and she looked at a proximate co-participant when she played the role of the other characters, the immigration officials.

I also demonstrated that the storyteller used gaze and pointing gestures to shift the roles of a co-participant from a listener to one of the story characters. When the storyteller looked at and pointed at the proximate co-participant with her right fingers and demonstrated one of the characters, the immigration official, she treated the co-participant as another character, Asai herself. She used her proximate co-participant as a story material in order to make the story lively, to depict the scenes more vividly, and to involve co-participants in her story. My analysis suggests the importance of analyzing conversational data from multiple perspectives including prosody, gaze, and body movements as well as syntactic structure in order to reveal a true picture of participation and involvement in a conversation.

**Appendix 1: Abbreviation of the gloss used in transcription**

|      |                  |     |                  |
|------|------------------|-----|------------------|
| AUX  | auxiliary        | QT  | quotative marker |
| PASS | passive morpheme | SUB | subject marker   |

**Appendix 2: Transcription conventions in transcription**

|             |                                    |       |                                 |
|-------------|------------------------------------|-------|---------------------------------|
| .           | falling intonation                 | [     | overlap                         |
| ,           | continuing intonation              | (0.5) | silence; in tenths of a second  |
| ?           | rising intonation                  | (( )) | nonverbal behavior              |
| :           | vowel lengthening                  | ↑↓    | sharper rises or falls in pitch |
| =           | “latched” utterances               | ><    | compressed utterances           |
| -           | a cut-off or self-interruption     | (h)   | laughter within a word          |
| <u>word</u> | emphasis by loudness or high pitch |       |                                 |

**References**

- Clark, Herbert and Richard Gerrig. 1990. Quotation as demonstration. *Language* 66:764-805.
- Goffman, Erving. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Goodwin, Charles. 1984. Notes on story structure and the organization of participation. In J. M. Atkinson and J. Heritage, eds., *Structures of social action: studies in conversation analysis*, 225-246. Cambridge: Cambridge University Press.
- Goodwin, Marjorie Harness. 1997. Byplay: Negotiating evaluation in storytelling. In G. R. Guy, C. Feagin, D. Schiffin and J. Baugh, eds., *Towards a social science of language: Papers in honor of William Labov Volume 2: Social interaction and discourse structures*, 77-102. Amsterdam: John Benjamins.
- Makino, Seiichi and Michio Tsutsui. 1986. *Nihongo kihon bunpo jiten* (A dictionary of basic Japanese grammar). Tokyo: The Japan Times.

290 Royce Hall Box 951540  
University of California, Los Angeles  
Los Angeles, CA 90095-1540

chisatok@ucla.edu

## **Gesture, Information Flow, and NP Form**

LORRAINE E. KUMPF

*California State University, Long Beach*

### **0. Introduction**

This paper attempts to join streams of research that have heretofore not been joined: those of information flow and gesture in classroom discourse. In this effort, I want to suggest the importance of genre in examining the relationships among pragmatics, grammar, and gesture. The term information flow is used as Chafe (e.g. 1994) uses it to refer to the changing status of information in ongoing talk, particularly the changing form and cognitive status of referents along the new/given continuum. In Section 2, I will report on results of work (Kumpf to appear) on the information flow characteristics of referents used by some teachers in American public high school science classes. This work correlates NP form and the given/new status of NPs and demonstrates some fundamental properties of NP distribution, clausal syntax, and information status in the teachers' discourse. It provides the basis for the consideration of gesture.

The analysis of the gestures used by the teachers constitutes the present study, in Section 3. The study is confined to their "deictic" (pointing) gestures. It will be demonstrated that these gestures are closely tracked with the flow of information. The general questions guiding this study are: How do the teachers use gesture to support the teaching of science content in the classroom? What kind of information accompanies deictic gestures, and what is its grammatical form?

The teachers' talk I sampled has grammatical and pragmatic traits which relate directly to their primary concerns in communicating the science content: teachers must make the content available to a large group (30-35 students); they must keep the students' attention focused on science content; and they must create the conditions for the students' retention of the content. This, they do in talk by the staging and highlighting of key items and processes. I will show that the teachers' deictic gestures interact with NP form and information status to support these classroom goals.

### **1. Previous research**

The literature on information flow is extensive. Chafe's research in this area (e.g. 1987, 1994) has included considerations of the cognitive status of informa-

tion from the perspective of what the speaker assumes the hearer to know or have in mind. Research influenced by Chafe integrates notions of the relative activation of speaker and hearer consciousness regarding the referent under discussion. He employs a three-part analytic system to represent the activation status of concepts: given information (concepts which the speaker assumes that the hearer has in active consciousness), new information (concepts which the speaker first brings to the hearer's consciousness), and accessible information (concepts which are available to the hearer through prior mention, inferencing, or evoking a frame). New information is typically expressed in full lexical nouns; given information may be attenuated (proforms or zeros in English), though a full range of structural possibilities can express givenness (Givon 1990, Ariel 1990, Chafe 1994). Another influential taxonomy for given/new, which does not include implications of consciousness, is offered by Prince (1981, 1992). Related studies on noun phrase accessibility (e.g. Ariel 1990) and on topic continuity in discourse (e.g. Givon 1983) lend perspective to this study.

Virtually all information flow research presupposes the interaction of a single speaker and hearer. In this study, it will be shown that some of the usual assumptions regarding information status and NP form may not apply to the genre of classroom discourse. Genre differences may thus call into question the way in which information flow analysis proceeds.

Analysis of classroom language is dominated by ethnographic approaches. Ethnographers have, for example, described recurring patterns such as the resilient formula "teacher initiates, student responds, teacher evaluates"—the "IRE sequence". Research on the language that teachers use in the classroom has often been concerned with speech acts, and especially the use of questions and directives (e.g. Heath 1982, Dillon 1990). Approaches addressing grammar or structural correlates to pragmatics in the classroom are scarce: Chaudron (1985), for one example, looks at the structure of topics in the classroom. There are apparently no studies on information flow in teachers' talk, excepting Kumpf (to appear), summarized below. A thorough review of classroom discourse is outside the scope of this paper; see Cazden (1988) for a summary.

Of the extensive literature on gesture, the work of MacNeill (e.g. 1992), Kendon (e.g. 1995), and Goodwin (1994, n.d.) have been most helpful in understanding deictic gestures. Gesture studies which incorporate information flow are rare. Kendon (1995) concludes that gesture falls on the new information in speech; this is a generalization that will be re-examined in the present study. Gesture in the classroom has been examined (summarized in Gullberg (1998)), but not in relation to information status.

In this study, it is assumed that the teachers' gestures and the characteristics of grammar and information flow in their language will relate directly to their goals in communicating science. Classroom research such as Cazden's (1988) define those goals: 1) to establish joint attention with the group; 2) to make information maximally accessible; and 3) to facilitate the retention of content.

## **2. Information flow in science teachers' talk: the background study**

In this section, I will summarize relevant results from a larger study of information flow in the classroom (Kumpf to appear). The purpose of this study was to link grammatical form and pragmatic status of NPs in classroom discourse. That is, NP form (lexical, pronominal, or zero) was correlated with the information status of the NP (new, given, or accessible), thus showing the typical argument patterning in the teachers' discourse. (See Du Bois (to appear) for general implications of these argument structure correlations.) Since teachers tended to build on the known, and since referents were accessible through their visible presence in the classroom, it was hypothesized that there would be few new mentions. A second hypothesis stated that there would be a high degree of lexical mentions in the discourse, regardless of their information status.

The data used in the analysis were four classes of tenth grade life science: two classes on genetics, one on electricity, and one on ecology. The students were 15-16 years of age, and there were 30-35 students per class. The teachers were experienced, recognized as competent, and taped at the suggestion of their principals. In the segments chosen for analysis, teachers spent most of the class time lecturing or reviewing material. That is, the classes were "teacher-fronted" and the teacher did almost all of the talking.

Relevant to the present study, NPs were coded for form (lexical, pronominal, zero), for grammatical role (subject of the intransitive, agentlike NP of the transitive, object, indirect object, and so on), and for the information status of the concepts the NP referred to: new, given or accessible). (See Kumpf (to appear) for other sampling and coding details.)

Results were that, as hypothesized, new information, in the form of new nominal referents, was quite rare in the data. Referents were given (active in the immediate context) or accessible (available through prior mention or through their visible presence in the speaking context). These were referents that the teacher assumed the students could identify. Second, also as hypothesized, the proportion of lexical mentions was very high; non-new nominals tended to be expressed as full lexical items. In fact, the discourse was very "nouny"; the proportion of lexical mentions was higher than in English conversation or narrative (Kaarkainen 1996; Kumpf 1992).

It was also shown that teachers used argument structures which allowed for the "staging" of referents, that is, the placement of referents in positions of salience in the clause. In accomplishing this staging, teachers favored a particular clause type using "low-content" transitive verbs (clauses which are transitive in surface syntax but do not express the prototype of agent/action/affected object). Of these verbs, HAVE when used as an existential ("On page 400 we have a diagram") was prominent, but FIND, SHOW, SEE, GOT, and others were used. These verbs are used to present, point out, display, emphasize—important pedagogical functions in the science class. It is not surprising that deictic gestures accompany this kind of talk.

The use of full lexical nouns rather than reduced forms relates to the teachers' goals of attention and retention. Here, models presented in research on information flow, which are limited largely to dyadic conversation, need recasting. Most of these models say that the speaker judges what the hearer knows or has in mind. However, the teacher (or presumably anyone addressing a large audience) cannot assume that any one hearer has a particular piece of information in mind. To ensure maximum accessibility of the NP, the teacher mentions it in full form. Also, the teacher aims for retention of lexical items; repeating them in full form supports this goal.

Thus this type of discourse, centered on demonstration, uses grammar to achieve the syntactic staging of nominals through the choice of verb and the use of low-content transitives, existentials, deictic expressions, and other marked syntax. Along with the morphosyntax of staging for show, teachers use many symbols and representations, usually written on the board, as well as authentic materials and visual aids. It is the presentational or demonstrational character of the discourse that to a great extent defines the use of gesture by the teachers. Deictic gesture, as we shall see, supports reference to particular concepts and relationships in the science material.

### **3. Gesture and NPs in the teachers' talk: indicated NPs**

It is obvious upon viewing the teachers in action that referent staging is achieved in gesture as well as talk. In fact, their physical behavior which supports the presentation of nominal referents is energetic and compelling.

In this study, I look at the characteristics of nominals which are expressed along with deictic gesture. I call these nominals "indicated NPs." The aim is to relate the gesture to the form and information status of those nominal referents. Deictic gesture is a term widely used to refer to pointing movements, especially employing the index finger. MacNeill (1992) includes in this category other means of pointing, and the use of manipulated objects for the purpose of indicating. It is assumed that the teachers use explicit gestures in order to bring objects or representations into the students' visual focus and to direct their attention.

#### **3.1. Indicated NPs**

In the lesson on electricity, the teacher, while explaining series circuits, walks to the board in the front of the room and touches the words "circuit" and "series," which are listed there, as he says them. "Series" and "circuit" are thus "indicated NPs." In order to be coded as an indicated NP, the teacher's verbalized lexical item must clearly accompany the gesture: the verbalization and the gesture must overlap in real time. The most typical gestural act occurring with indicated NPs in the data is pointing to a representation on a whiteboard, usually in the front of the classroom. Figure 1 is illustrative. In the caption, square brackets indicate the point of overlap of the teacher's words and her deictic gesture.

*Figure 1.*



*Teacher:* what is ^in ... [those sperm and egg cells.]  
[T INDICATES DRAWING ON BOARD]

### **3.2. Research Questions**

1. What type of referent is an indicated NP, in terms of its scope of reference (i.e., specific, generic, identifiable, etc.)?
2. What is the information status of indicated NPs?
3. What is the grammatical form of indicated NPs?

### **3.3. Methodology**

The data for this study is a subset of the science class data used for the information flow analysis summarized above. Three of the teachers' tapes were analyzed. The transcripts were amended to note the gesture; sets of square brackets indicate the overlap with talk. Each teacher's data were sampled until 50 occurrences of indicated NPs were recorded. The teachers' gestures that are not deictic, and do not point to particular nominal referents, were not analyzed.

For all NPs (indicated or not), the following points were coded: 1) scope of reference of NP (general, generic, identifiable, exemplar, etc.); 2) information status of the concept expressed in the NP: given, new, or accessible. The information status was arrived at by textual analysis. A concept was coded *given* when continuous in reference. First and second person referents were assumed given. A concept was *new* when first mentioned. A concept was *accessible* by prior mention, with the arbitrary limit of 20 clauses prior, or if the referent was in the physical environment, visible to speaker and audience, or if the referent could be inferred on the basis of shared frames or general knowledge. (This last was rare in these data.); 3) Grammatical form of indicated NP. NPs were found to belong to the following categories: indefinite lexical NP, demonstrative with lexical NP, definite article with lexical NP, possessive with lexical NP, demonstrative pronoun, unstressed pronoun, and other.

### 3.4. Results

#### 3.4.1. Teachers' deictic gestures used in indicating NPs

Gestures typically take place at the board. The teachers:

- point with the index finger at a representation or object, especially using the whiteboard or chalkboard.
- touch, underline use a pointer or otherwise indicate an NP that is on the board, including slapping or hitting the board, or using sawing motions under the representation.
- trace, usually with the index finger, a line, representation or figure.
- pick up an object or representation and hold it for the class to see, while pointing.
- use other movements or body parts to indicate a nominal.

All of these actions are considered to be in the class of deictic gestures in McNeill (1992), Goodwin (1994), and others. I noted that, while talking, teachers would sometimes write illegibly or draw uninterpretable representations on the board in movements which were gestural and probably deictic, but these, as well as some other actions that were difficult to interpret, were not included in this study.

#### 3.4.2. Three explicated illustrations

In Figure 2, the teacher is explaining a formula on the board and is naming an element of the formula (“amperage”) while pointing to its representation (I). This is part of a longer presentation, in which he names the terms repeatedly, while pointing to the relevant symbols. Several times, he reiterates the combined action of pointing directly under the symbol in the formula and naming its referent. In addition to the referents, he thus expresses the relationships of the concepts in the formula. These concepts challenge the students; hence, the explicit repetition. With his actions, the teacher is also addressing the goal of student retention.

*Figure 2.*



Teacher: Power equals ^voltage times [^amperage.]  
[T TOUCHES BOARD UNDER LETTER 'I']

*Gesture, Information Flow, and NP Form*

The next example is meant to illustrate that the deictic “point” may also be abstract, that is, may point “to” a concept that is not present, to an abstraction that is not concretized in a representation, or to a conceptual space. Such a gesture may be accompanied by an indicated NP. MacNeill (1992) and others have shown these gestures to reflect the thought process of the speaker, but I suggest that in the classroom, they may also act as “pointers” to the hearer. In Figure 3, the teacher is specifying a value of the formula that he wants the students to find. As he points “to” the abstract NP as a focus of attention, the downward trajectory of the teacher’s index finger stops with stressed syllable of the NP.

*Figure 3.*



Teacher: I want you to look for the [<sup>^</sup>voltage].  
[T’S INDEX FINGER STOPS  
ITS DOWNWARD TRAJECTORY]

In the final example, the teacher is talking about the joining of male and female chromosomes. Figure 4 shows her using objects (models of chromosomes)

*Figure 4.*



Teacher: ...with <sup>^</sup>her [chromosomes]  
[T PUTS MODEL CHOMOSOMES ON BOARD]



straightforward way, e.g. “Here we see chromosomes.” The concepts may also be encoded in other ways, as in “I have my power source,” where the teacher means any power source. This appears to be a basic strategy in teachers’ science talk

An analysis of *non-indicated* NPs is not included here; however, a preliminary analysis of such NPs, excluding first and second person references, shows that about one-fourth of non-indicated NPs are exemplars. Put another way, the great majority of NPs that are mentioned without indication are not exemplars of a class of nouns. If verified, this result shows that teachers use the gestures significantly less often with non-exemplars. Regardless, the data in Table 1 demonstrate that teachers show a preference for using deictic gestures with NPs representing the key science content.

**3.4.4. What is the information status of indicated NPs?**

Table 2 shows that most NPs that are indicated are also given or accessible. The teachers are pointing to representations that have been mentioned or are visible in the context.

|              | Given<br>N (%) | Accessible<br>N (%) | New<br>N (%) | Total      |
|--------------|----------------|---------------------|--------------|------------|
| Teacher A    | 35(70)         | 15(30)              | 0            | 50         |
| Teacher B    | 42(84)         | 7(14)               | 1(2)         | 50         |
| Teacher C    | 28(56)         | 21(42)              | 1(2)         | 50         |
| <b>Total</b> | <b>105(70)</b> | <b>43(29)</b>       | <b>2(1)</b>  | <b>150</b> |

*Table 2.* Information Status of Indicated NPs

Regarding the rest of the data, the information status of non-indicated NPs distributes comparably: given information is 74%; accessible, 20%; and new, 6%. Thus, as was mentioned in section 2 above, there is very little new information in the discourse, in the form of new mentions.

**3.4.5. What is the grammatical form of indicated NPs?**

Looking at the forms in Table 3, we see that, of the NP forms which occur in the data, the largest number of indicated NPs is contained in the category Indefinite Lexical NP, at 40% of the total. Though indefinites are commonly used, we note that the other form categories represent definite NPs, and taken together, 60% of the indicated NPs are definite. One might assume indicated NPs to be expressed as definite, since they are identified. However, the picture is more complex, and the nature of these referents needs further specification.

|              | Indef<br>LNP<br>N (%) | Demon<br>+LNP<br>N (%) | Poss<br>+LNP<br>N (%) | Def Art<br>+LNP<br>N (%) | Demon<br>PN<br>N (%) | Unstr<br>PN<br>N (%) | Other<br>N (%) | Total      |
|--------------|-----------------------|------------------------|-----------------------|--------------------------|----------------------|----------------------|----------------|------------|
| Teacher A    | 17(34)                | 5(10)                  | 10(20)                | 4(8)                     | 6(12)                | 8(16)                | 0              | 50         |
| Teacher B    | 26(52)                | 5(10)                  | 3(6)                  | 5(10)                    | 8(16)                | 0                    | 3(6)           | 50         |
| Teacher C    | 18(40)                | 9(18)                  | 5(10)                 | 8(16)                    | 6(12)                | 1(2)                 | 3(6)           | 50         |
| <b>Total</b> | <b>61(40)</b>         | <b>19(13)</b>          | <b>18(12)</b>         | <b>17(11)</b>            | <b>20(13)</b>        | <b>9(6)</b>          | <b>6(4)</b>    | <b>150</b> |

Table 3. Grammatical form of indicated NPs

Table 3 also reflects the general finding that indicated NPs are lexical. They occur as lexical in indefinites, with a demonstrative, with a definite article, and with possessives. In total, seventy-six percent of indicated NPs are lexical.

#### 4. Discussion

Many discourse analysts have made claims about the relative lightness of given or identifiable information. For example, Ariel (1990) claims that the most attenuated NP possible for the context will be the preferred one. But the teachers violate this, in light of their goals of student attention and retention. Although they are constantly monitoring for attention, the teachers presumably cannot be sure that any one student will be focused on the referent. Thus, the NP is repeated in full form, and likely with gestural support. The teachers monitor the attention of students and modify their strategies “on line.”

Retention in the complex classroom context is a challenging goal. Since pronouns are not memorable, lexical NPs are repeated. Gesture also supports retention. In his research in this area, Pavio (1986) refers to “dual coding systems,” the verbal and the visual. When material is coded in both systems in one communicative context, memory for the material is strongest.

As suggested above, many of the findings regarding information flow have been established by studies of oral narratives or dyadic conversation; these findings cannot easily be applied to classroom language. In the literature, it is claimed that referents are accessible and identifiable when present in the context, but co-presence in the classroom is not enough. It is also necessary to get that referent into play—to get it from “the context” into shared consciousness. Clearly, deictic gesture is vitally important in what I have called the staging of the referent.

Getting the referent into play is a much greater challenge for a teacher than for speakers in a conversation or narrative. Besides the factors of student attention and size of audience, the type of referent needs to be considered. In the science classes teachers talk about abstract concepts and make general references. They talk in language that is far removed from the daily life of students, and they make relationships between abstract referents which are challenging to understand. It is not surprising that teachers encode material in as many ways as they can.

## **5. Conclusion**

Whereas this study has shown some typical relationships between gesture, NP forms and the information status of nominals in the classroom, there are a number of limitations. Gesture is continual in the teachers' presentations, and many other types of gestures are present. I have severely limited this examination of gesture and NPs in order to express one small aspect of a very complex communicative situation. By insisting that the NP vocalization be synchronized with the gesture, I eliminated many other gestures that may be interpreted as deictic, and may support science content. Also, by limiting the study to deixis, I have ignored possible ways in which other gestures support the information.

Complex gestural behavior may serve many functions. For example, often a gesture that is iconic may also have a deictic function. (MacNeill (1992) discusses this multi-functionality.) Again, such complexity was avoided by limiting this study. Similarly, I set aside problems of the scope of the gesture vis a vis the accompanying talk, by insisting on the clear referentiality of the pointing behavior. A gesture may relate more approximately to the talk, for example, by relating to a complex of ideas or a series of clauses. Also, a given deictic gesture may not be simultaneous with the vocalization. Furthermore, additional ways in which gesture can relate to the talk were unexamined. Some deixis, for example, had no speech accompaniment, even though the referent was clear. Thus the narrowness of the present study leaves much of the teachers' gestural behavior to be examined. These limitations notwithstanding, the study has begun to look at the ways in which gesture, status of information and grammatical form work together to realize the management of information in the classroom. The norms of the genre define the specific shape of this achievement.

## **References**

- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Cazden, Courtney. 1988. *Classroom Discourse*. Portsmouth, NH: Heinle & Heinle.
- Chafe, Wallace. 1994. *Discourse, consciousness and time*. Chicago: University of Chicago Press.
- Chafe, Wallace. 1987. Constraints on information flow. In R. Tomlin, ed., *Coherence and grounding in discourse*. Amsterdam: John Benjamins.
- Dillon, J. T. 1990. *The practice of questioning*. London: Routledge.
- Du Bois, John. To appear. Discourse and grammar. In M. Tomasello, ed., *The New psychology of language: Cognitive and functional approaches to language structure, Volume 2*. Hillsdale, NJ: Lawrence Erlbaum.
- Du Bois, John, Lorraine Kumpf, and William Ashby, eds. To appear. *Preferred argument structure: Grammar as architecture for function*. Amsterdam: John Benjamins.

Lorraine E. Kumpf

- Du Bois, John, et al., eds. 1993. Outline of discourse transcription. In J. Edwards and M. Lampert, eds., *Talking data: Transcription and coding in discourse research*. Hillsdale, NJ: Lawrence Erlbaum.
- Givon, Talmy. 1990. *Syntax: A functional-typological introduction, Volume 2*. Amsterdam: John Benjamins.
- Givon, Talmy, ed. 1983. *Topic continuity in discourse*. Amsterdam: John Benjamins.
- Goodwin, Charles. 1994. Professional vision. *American Anthropologist* 96(3): 606-33.
- Goodwin, Charles. N.d. Pointing as situated practice. Ms.
- Gullberg, Marianne. 1998. *Gesture as a communication strategy in second language discourse*. Lund: Lund University Press.
- Heath, Shirley. 1982. Questioning at home and school. In G. Spindler, ed., *Doing the ethnography of schooling*. New York: Holt, Rinehart and Winston.
- Kendon, Adam. 1995. Gestures as illocutionary and discourse structure markers. *Journal of Pragmatics* 23(3):247-79.
- Karkkainen, Elise. 1996. Preferred argument structure and the subject role in American English conversational discourse. *Journal of Pragmatics* 25(5): 675-701.
- Kumpf, Lorraine. To appear. Genre and preferred argument structure. In J. Du Bois, L. Kumpf, and W. Ashby, eds., *Preferred argument structure: Grammar as architecture for function*. Amsterdam: John Benjamins.
- Kumpf, Lorraine. 1992. Preferred argument structure and second language discourse. In *Studies in Language* 16(2): 369-403.
- McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, David, ed. 2000. *Language and gesture*. Cambridge: Cambridge University Press.
- Pavio, A. 1986. *Mental representations: a dual coding approach*. Oxford: Oxford University Press.
- Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness, and information status. In W. Mann and S. Thompson, eds., *Discourse description*. Amsterdam: Benjamins.
- Prince, Ellen. 1981. Toward a taxonomy of given-new information. In P. Cole, ed., *Radical Pragmatics*. New York: Academic Press.

Department of Linguistics  
California State University, Long Beach  
Long Beach, CA 90840

lekumpf@csulb.edu

## Grammar and Gesture in American Sign Language: Implications for Constructing Meaning

SCOTT K. LIDDELL  
*Gallaudet University*

### 0. Introduction

Stokoe (1960) was the first to argue that American Sign Language (ASL) demonstrates the kind of abstract structure one expects to find in the analysis of any spoken language.<sup>1</sup> The idea that ASL was not a real language was so deeply embedded in the culture, however, that it required nearly two decades for Stokoe's claim to gain acceptance. For Stokoe's idea to take hold, it was necessary to give up the idea that ASL was a poor *gestural* substitute for real language. Linguists analyzing ASL found morphemes where previously there had been only gestures. The transformation from gesture to morpheme was so complete that, in the end, no gestures remained.

The fact that some signs point toward things is undeniable. ASL pronouns, for example, point toward physically present referents. The pointing, however, has not become part of the analysis of these pointing signs. The field has come to accept that even when a referent is physically present, a sign is articulated at an area of space associated with the physically present referent. Claiming that signs are articulated with respect to locations in space avoids the necessity of claiming that signs actually point at things. The analysis is the same for non-present referents, where an area of space is also associated with the non-present referent. In either case, signs are seen as being directed toward morphemic areas of space.

Liddell (1995) argues that if the referent is physically present, signs are not directed toward areas of space at all, but rather, toward the referent. In addition, I analyze areas of space associated with non-present referents as conceptual entities rather than morphemes. Furthermore, signs point at such areas of space because of the conceptual ability to point. This challenges the notion that signs point because the signer is articulating a spatial morpheme. There has been considerable opposition to the idea that what looks like pointing really is pointing. Part of the reason

---

<sup>1</sup> I would like to thank MJ Bienvenu, Melissa Draganac, Paul Dudis, and Greg Visco for very useful discussions of the ASL data and for providing their native speaker intuitions about the use of directional verbs.

for this is that it gives the appearance of a return to the pre-Stokoe idea that ASL was merely a collection of gestures rather than a language. A second reason for the opposition rests with assumptions within the field of linguistics itself. Ideas about the structure of ASL were being transformed during the seventies and eighties. The operating principles during those times held that the meaning of a sentence comes from its morphemic parts and their grammatical arrangement. Langacker (1991) refers to this as the “building block metaphor.” Since directing signs toward things is clearly meaningful, there was no choice but to make morphemes responsible for both the meaning and the directionality.

### **1. Directional Pronouns in ASL**

There is a significant articulatory difference between all sign languages and all spoken languages. In producing spoken words, the tongue does not point at things in the environment as a speaker produces words. In contrast to this apparently universal characteristic of speech, several classes of signs do point at things. That is, in order to properly produce many ASL signs, the hands *must* be directed toward real people or things being talked about, if those real people or things are present. All known sign languages include such signs.

During the past thirty years, there have been numerous proposed morphemic analyses of directional signs including Woodward (1970), Lacy (1974), Fischer (1975), Friedman (1975), Edge and Herrmann (1977), Klima and Bellugi (1979), Gee and Kegl (1982), Padden (1988), and Neidle et al. (1995). All such proposals are based on the idea that morphemic locations in the space ahead of the signer become associated with a referent—whether physically present or not. The analyses assume that incorporating a spatial locus into the structure of the verb causes the spatial locus to become the point of articulation for the verb. This causes the hand as articulator to move toward that locus.

Liddell (1994, 1995) demonstrates that the assumption upon which these analyses are based is false. That is, signs do not move toward a single spatial locus. Analyses based on the idea that signs do move this way are inconsistent with observable facts about how signs are actually produced. I have argued that pronouns and directional verbs either point at physically present referents themselves or point at conceptualized-as-present referents. I will not reiterate the arguments here, but rather, will take it as given that the types of directional signs discussed in this paper actually point at things. Sign types that can be gradiently placed or directed for the purpose of pointing include some nouns, almost all pronouns, a large set of verbs, numeral quantifiers in general, signs expressing locative meanings, and even signs for alphabetic characters.

Signs can point by the directionality of their movement. The way the hand is oriented can also point. For example, in Figure 1 the singular, non-first person pronoun PRO has the index finger directed toward and also moves toward the addressee. In Figure 1a, PRO is directed horizontally toward the addressee’s chest. In Figure 1b, PRO is also directed toward the addressee’s chest, but its form differs from the sign in Figure 1a. The reason for this is straightforward. Since the

referent is standing in Figure 1b, PRO must be directed upward in order to point at his chest.

**Figure 1**

a. PRO



b. PRO



In Figure 2a, the dual first person pronoun makes an in-and-out rocking motion between the signer and the other person it is directed toward. The motion is primarily due to bending the wrist toward and then away from the signer. The in-and-out motion produces a line whose two ends point at the two referents of the pronoun. Motion of the wrist also causes the non-first person dual pronoun in Figure 2b to make a sideways back-and-forth movement that also points toward the pronoun's two referents.

**Figure 2**

a. PRO-1-DUAL



b. PRO-DUAL



## **2. Accounting for Directionality**

The signs in Figures 1 and 2 illustrate the general point that ASL pronouns have fixed lexical meanings and derive their contextually variable referential values from their directionality. The referential values are indicated by the directionality of the signs themselves. This fact about ASL is easily accommodated using concepts from cognitive grammar (Langacker 1987, 1991) and mental space theory (Fauconnier 1985, 1997).

In cognitive grammar a noun or pronoun profiles a thing in some domain. This constitutes the lexically encoded meaning of the noun or pronoun. Figure 3 contains two representations of the semantic structure of PRO. In Figure 3a, the semantic structure includes two entities: the signer, represented as a circle with an

'S' inside it, and one additional entity. The second entity, represented as a bold circle, is the profiled entity in the semantic structure of PRO. This is a diagrammatic way of representing some entity other than the signer. An abbreviated representation of the same semantic structure that only represents the profiled entity itself appears in Figure 3b. I will use this type of abbreviated representation in subsequent diagrams.

**Figure 3**



Liddell (1995) proposes that one's conception of the immediate environment is a mental space called *real space*. Real space differs from other mental spaces in that the entities within it are conceptualized as existing in the immediate environment itself. In general, our conceptualizations of immediate reality conform well to the actual physical elements we assume are responsible for our perceptions. That is, I take in perceptions, and construct real space based on those perceptions. If my real space includes a pencil on a desk in front of me, then I have confidence that if I extend my hand toward the real space pencil, I will make contact with a physical pencil. Instances where real space does not conform to the actual locations of physical entities help show the distinction between conceptual entities in real space and physical things. For example, Figure 4 is a picture of an aquarium. The camera was placed so that the corner of the aquarium appears in the center of the picture.

**Figure 4**



From this angle there appear to be four fish in the aquarium. In fact, there are only three. The two fish labeled 'Fish #1' appear both to the left and to the right of the corner of the aquarium. For a person perceiving the aquarium from the same vantage point as the camera, real space contains four fish, even though the

aquarium only contains three. Not only does real space contain four fish, the locations of the real space fish do not conform to the locations of the actual physical fish. This is obvious since Fish #1 appears to be in two different places. It does not even make sense to ask which one is the actual physical fish since neither one of the real space fish is in a location corresponding to the location of the physical fish.

In our daily lives, we move around and interact with our environment based on real space—our mental space representation of our environment. In general, the locations of the conceptual entities in real space correspond well to the locations of the physical things we interact with. The correspondence is so good, in fact, that we do not distinguish real space from reality. We operate as if real space were reality.

In mental space theory, meaning is constructed by making connections between entities within mental spaces. Fauconnier (1985, 1997) demonstrates that such mappings are a crucial, ongoing, and constant aspect of the use of any language. Thus, we would expect that mental space mappings also play a crucial role in constructing meaning in ASL.

The directionality of pronouns and indicating verbs is best understood as providing mental space mapping instructions. For example, suppose a signer is facing her addressee and directs the singular non-first person pronoun PRO toward a man to her right and to the left of the addressee. The addressee will recognize the sign PRO, which will lead the addressee to include the semantic structure of PRO (“one entity other than the signer”) as part of the semantic structure of the utterance. Not only does the addressee recognize her production as an instance of PRO, the addressee also sees that PRO’s directionality leads to the man to the left of the addressee.

**Figure 5**

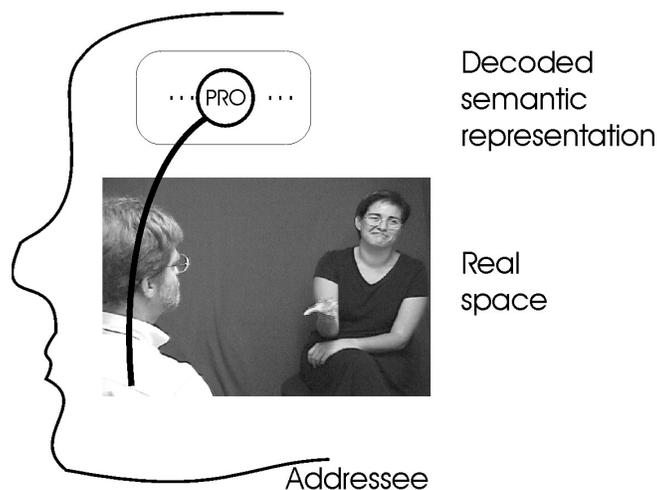


Figure 5 represents PRO as part of the semantic structure of the utterance being produced and also represents the addressee’s real space. The significance of

PRO's directionality is that it provides an instruction to make a mental space mapping between the semantic structure of PRO and the entity it is directed toward. Since PRO is directed toward the man to the left in the addressee's real space, PRO maps onto that man.

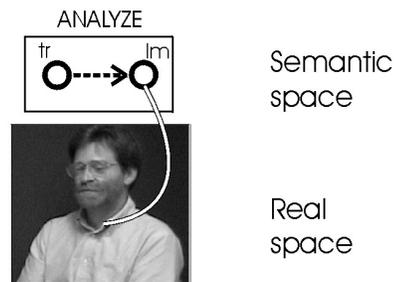
The directionality of indicating verbs is also best understood as providing mental space mapping instructions. In Figure 6a, for example, the sign ANALYZE is directed toward the addressee, directly ahead of the signer. In a cognitive grammar semantic representation, a verb like ANALYZE contains a trajector and a landmark. A trajector is the primary figure within a profiled relationship while the landmark is the secondary figure (Langacker 1987, 1991). In the active verb ANALYZE, the primary figure is the entity carrying out the analysis and the landmark is the entity being analyzed. In general, an indicating verb moves toward the entity to be mapped onto its landmark.<sup>2</sup> Since ANALYZE is directed toward the addressee in real space, this prompts a mapping between him and the landmark, shown by the connector between the two in Figure 6b.

**Figure 6**

a. ANALYZE



b. Mapped semantic structure



For verbs with a plural landmark, the directionality often takes the form of a sweeping motion that moves past the multiple entities, thereby indicating the extent of the group of entities. For example, Figure 7 illustrates two instances of COMPARE<sup>[exhaustive]</sup>.<sup>3</sup> In Figure 7a, the hands move past a group of videotapes on a table in front of the signer. The sweeping movement of the hands indicates the entities to map onto the verb's landmark.

<sup>2</sup> There are some exceptional indicating verbs that move away from the entity to be mapped onto the landmark. Using a different analytical framework, Padden (1988) refers to them as “backwards verbs.”

<sup>3</sup> Klima and Bellugi (1979) treat the sign shown in Figure 7a as a ‘seriated external’ inflection of the verb COMPARE. They analyze a vertically downward moving form of the sign as a ‘seriated internal’ inflection of COMPARE. In their analysis the difference in meaning between the two forms is inflectional. In the analysis I am describing in the text the difference in meaning between the two signs illustrated is not an inflectional difference but a referential one, based on the directionality of the sign.

Another instance of COMPARE<sup>[exhaustive]</sup> is illustrated in Figure 7b. In this case the verb makes a diagonally downward sweeping motion, indicating the diagonally arranged set of pictures.

**Figure 7**

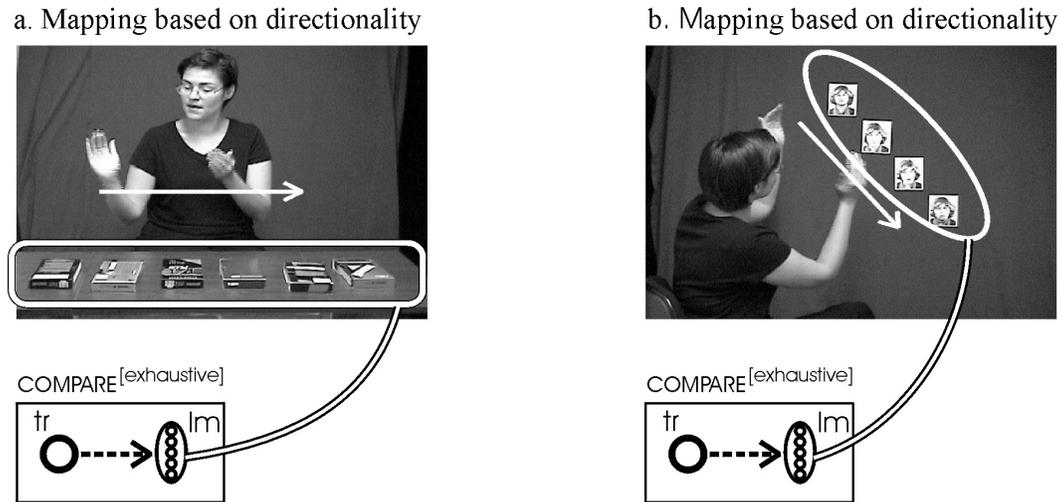


Figure 7a illustrates the mapping between the semantic structure of the verb and the set of videotapes indicated by the horizontal sweep of the hands. The mapping is motivated solely by the directional sweeping movement of the hands as the verb is produced. Figure 7b shows the mapping that results from directing the same verb along a diagonal path past the pictures in Figure 7b. The same semantic structure maps onto a distinct set of entities.

### 3. Mappings beyond the Individual Sign

Thus far I have restricted the discussion of directionality to individual signs. I have attempted to demonstrate that the directionality of both pronouns and indicating verbs is best understood as identifying one or more entities to be mapped onto the semantic structure of the sign itself. Next I will describe what happens when a verb and its pronominal subject both point.

**Figure 8**

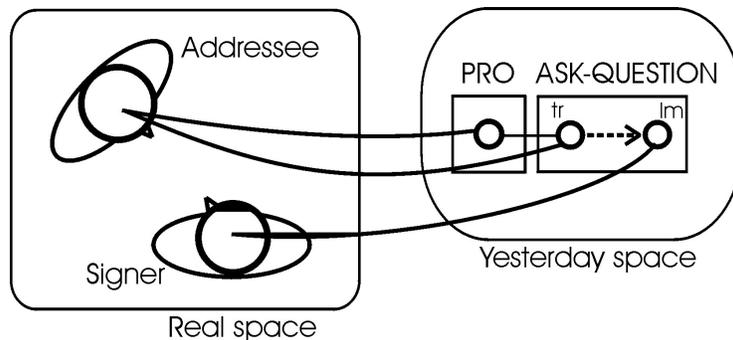


Figure 8 illustrates the mapped semantic structure of the clause PRO ASK-QUESTION YESTERDAY. In this instance, the subject, PRO, is directed toward the addressee. The verb ASK-QUESTION begins its movement directed toward the addressee then moves in a straight path toward the signer.

Following Langacker (1999), I treat semantic structures as mental spaces. Thus, I am representing the semantic structure of the three-sign clause PRO ASK-QUESTION YESTERDAY as a mental space with the setting ‘yesterday’. Because of the subject-verb relationship between PRO and ASK-QUESTION, the semantic structure of the pronoun and the verb’s trajector are linked. The horizontal connector between those two conceptual entities represents this connection.

Directing the subject pronoun PRO toward the addressee is an instruction to map the real space addressee onto the entity profiled in the semantic structure of the pronoun. Beginning the verb ASK-QUESTION toward the addressee is an instruction to map the addressee onto its trajector. Moving the verb along a path toward the signer is an instruction to map the landmark onto the signer. The resulting set of mappings involves two mental spaces: the semantic representation of the clause and real space. Simply following the mapping instructions from the directionality of the individual signs creates a coherent representation of the constructed meaning of this utterance.

This example is typical in that the directionality of a subject pronoun and the verb provide mapping instructions that result in the subject pronoun and the trajector being mapped onto the same entity—in this case, the addressee in real space. This does not always happen, however, as will become apparent later.

#### **4. Making Entities Present through Blending**

In the next example the signer is discussing the number of minutes in college and professional basketball games. He begins with the sign BASKETBALL, marked as a topic by raising the eyebrows and tilting his head back (Liddell 1980).

Next he makes college and professional basketball present in the space ahead of him by producing the sign COLLEGE to the left of center and #PRO (‘professional’) to the right of center, as illustrated in the second and third frames of Figure 9. The entities ‘college basketball’ and ‘professional basketball’ are quite abstract. Nevertheless, mental space blending (Fauconnier and Turner 1996, Fauconnier 1997) allows the signer to make those abstract entities present.

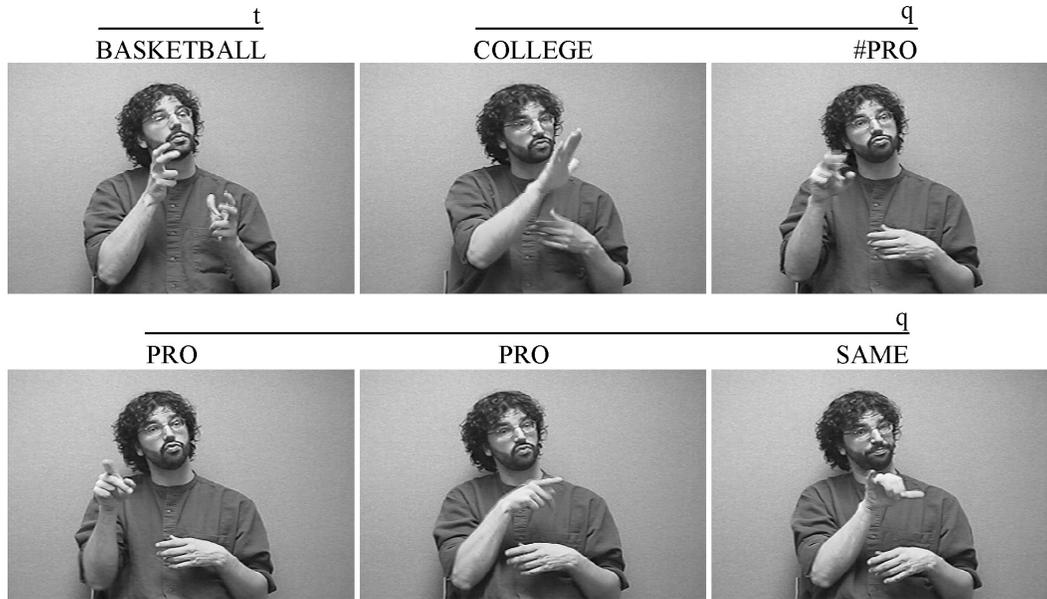
Next he directs one instance of PRO toward the area of space blended with professional basketball, which I label |professional basketball|, and directs a second instance of PRO toward the area of space labeled |college basketball|. The two instances of PRO serve as the dual subject of the side-to-side moving verb SAME. This particular instance of SAME moves back and forth between |college basketball| and |professional basketball|.

The directionality of the two pronouns and the verb in this example work just like the previous example where the actual referents were physically present. In this case, the abstract entities ‘college basketball’ and ‘professional basketball’ were made to be present through blending with the space ahead of the signer.

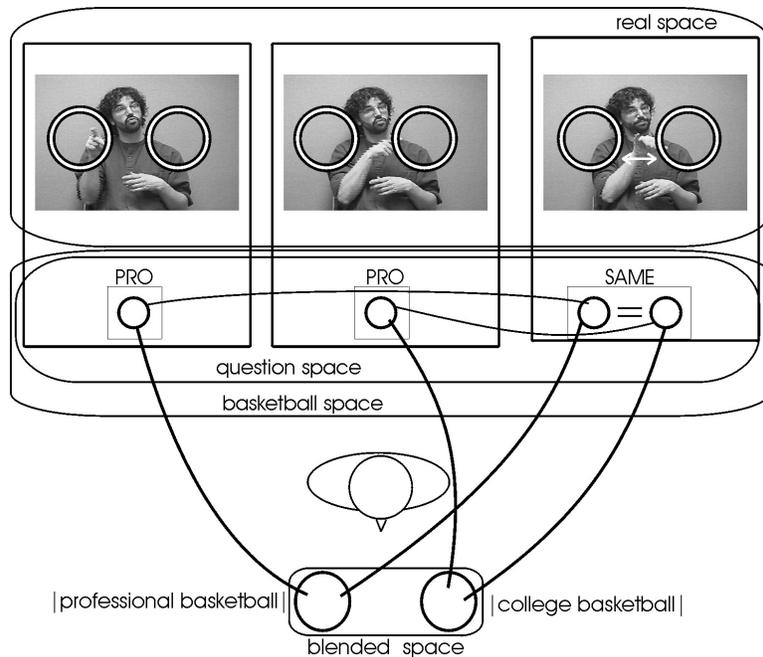
*Grammar and Gesture in American Sign Language*

Since blending made them present, both of the pronouns and the verb can be directed toward them. The mappings associated with the final three signs in Figure 9 are illustrated in Figure 10.

**Figure 9** “As for basketball, are college and professional games the same (with respect to their duration).”



**Figure 10**





In order to understand this example, it is necessary to describe the complex conceptualizations that underlie its use. First, the signer is describing an event that took place yesterday, represented as the ‘event space’ in the upper right in Figure 11. The sign LOOK-AT<sup>[durational]</sup> is being produced in real space, upper left in Figure 11. Through mental space blending, she projects aspects of the event space onto real space. The signer projects herself from the ‘event space’ onto herself in real space. I have labeled the resulting blended entity as the |signer|. <sup>4</sup> The person staring at her is also projected to the right of the |signer| in the blended space. I have labeled this entity as the |gazer|.

Through blending, the signer has recreated the scene in which the current addressee was staring at the side of her head yesterday. Only part of the scene is visible. The addressee can see the |signer| but cannot see the |gazer|. Because of the existence of the real space blend, there are two ‘realities’ that signs can be directed toward. One is real space and the other is the real space blend. An additional characteristic of the blended space adds to the complexity of this example. The blended space exists in the same place as real space. A non-signing passerby only has access to real space. The addressee, however, not only has access to real space, but also understands that aspects of the event that took place yesterday are also visible in the blend. Another interesting aspect of the blended space is that it has two distinct settings. It inherits its conceptual setting from the ‘yesterday’ event space. It inherits its physical setting from real space. Thus, it is a here-and-now partial recreation of the event that took place yesterday.

Given the existence of these two overlapping spaces, we can now explore the function of the pointing seen in both the subject and the verb. I will begin with the verb. During the production of LOOK-AT<sup>[durational]</sup> the back of the hand indicates that the |gazer| was doing the looking. The directionality of the fingertips indicates that the looking was directed toward the side of the head of the |signer|. If we make the assumption that the subject of the verb identifies the entity doing the looking, then the pointing of the subject pronoun PRO indicates that the current addressee was the one doing the looking. The result is a conceptually coherent set of mappings. The verb’s directionality identifies the two participants in the real space blend. The verb’s directionality does not, however, provide the identity of the |gazer|. The identity of the |gazer| only becomes apparent by directing the subject PRO toward the real space addressee.

## **5. Implications for Meaning Construction**

Currently, meaning construction is conceived of as taking place at a level distinct from language structure:

---

<sup>4</sup> The signer only partially becomes herself yesterday. That is, it is primarily her head that represents herself yesterday. The signer is still using her right hand in real space to narrate.

Meaning construction takes place at a “cognitive” level, call it *level C*; this level is distinct from the language structure (i.e. it is not an “underlying form,” it is not a “representation” of language or of language meaning, it is not objectively associated with any particular set of linguistic expressions). (Fauconnier 1997:36)

I have not been treating meaning construction as something that takes place at a level distinct from language structure here since the directional ASL signs appear to require mappings between their individual semantic structures and other mental space entities. Although it might be possible that sign languages and vocally produced languages operate differently with respect to meaning construction, it could also be possible that constructing meaning by mapping semantically encoded meaning onto other mental space elements can adequately account for meaning construction in both signed languages and vocally produced languages.

It is not possible to provide a thorough account here of how this approach to meaning construction can account for the wide array of data for which mental space theory has already demonstrated revealing solutions. However, below I will describe how this approach can account for the role-value distinction. I will then suggest an approach to metonymy, which also differs from the current mental space approach to metonymy.

I will use the sentence *Bob’s neighbor bought an island* to illustrate the distinction between roles and values.

**Figure 12**

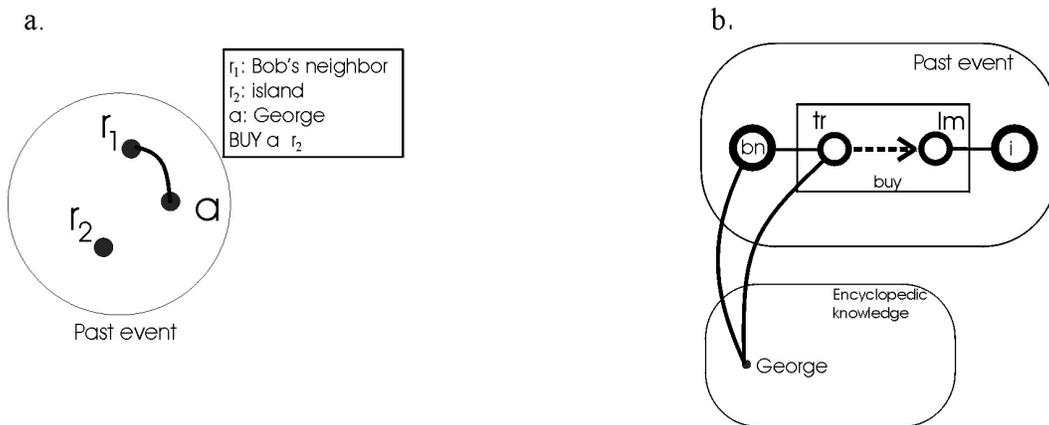


Figure 12a contains a mental space structured according the proposals in Fauconnier (1997). Entities  $r_1$  and  $r_2$  represent the two roles ‘Bob’s neighbor’ and ‘an island’. Assuming that the addressee knows that the speaker is talking about his neighbor George, entity  $r_1$  is mapped onto entity  $a$  (George). Entity  $r_2$  is unmapped and remains a role, unmapped onto a specific value.

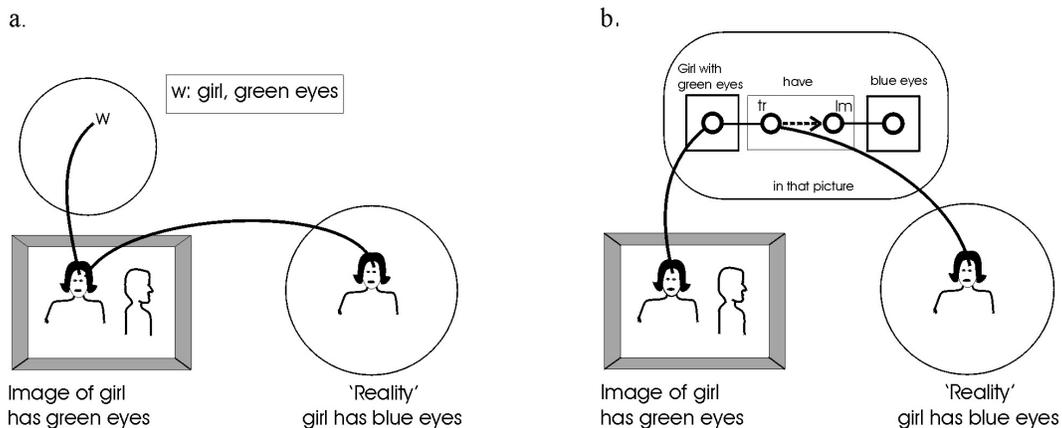
In Figure 12b *Bob’s neighbor* maps onto the entity *George* in the addressee’s encyclopedic knowledge. Since there is no metonymy in this example, the trajectory likewise maps onto George. However, the addressee does not know which

island is being talked about. In this circumstance, the semantic structure of *an island* remains unmapped.

A comparison of the two diagrams shows that the role  $r_1$  in Figure 12a corresponds to the semantic structure of *Bob's neighbor* in Figure 12b. Similarly,  $r_2$  corresponds to the semantic structure of *an island*. A conceptual role is an entity with a specific semantic property. That is, the role 'an island' is a conceptual entity with the property 'island'. This is exactly how one would describe the semantic structure of the noun phrase *an island*. As a result, the semantic poles of the subject and object noun phrases above could be described as expressing roles. Since *Bob's neighbor* maps onto George, the subject NP has a value reading. Thus, without the need to propose role-type entities, semantic representations automatically provide roles. When mapped onto other mental space entities, they express value readings.

Metonymy describes the use of language where a speaker mentions one thing in order to talk about another. Fauconnier (1985) proposes a solution to metonymy in which, given the appropriate pragmatic circumstances, mentioning one entity *a* can lead directly to a related entity *b*. Both *a* and *b* are mental space elements linked by a connector. Element *a* is called the trigger and element *b* is called the target. The ambiguous example *In that picture, the girl with green eyes has blue eyes* can be understood to mean that the image has green eyes while the model has blue eyes, or vice versa.<sup>5</sup>

Figure 13



The mapping shown in Figure 13a illustrates the constructed meaning in which the image of the girl has green eyes while the model has blue eyes. Entity *w* maps onto the image of the girl (the trigger), which does have green eyes. The image maps onto the model (the target), who has blue eyes.

If semantic spaces of the type proposed by Langacker (1987, 1991) are used for meaning construction, an alternative suggests itself. The directionality of ASL

<sup>5</sup> This example is similar to an example from Fauconnier (1985).

pronouns and verbs has already demonstrated that overt subjects and trajectors can be independently mapped in the process of meaning construction. In metonymy a person mentions one thing in order to talk about another. The solution proposed here is a direct representation of that description of metonymy. The subject *the girl with green eyes* maps in a straightforward way onto the image of a girl with green eyes. If the semantic structure and the mapping of an overt subject is taken as a guide in determining the mapping of the trajector, then the addressee is free to conclude that the entity with blue eyes is, in fact, the model. This allows a mapping between the trajector and the girl with blue eyes. Subjects and trajectors will typically map onto the same entity, but not in the case of metonymy, where the speaker mentions one thing in order to talk about another.

## 6. Conclusion

The directionality of pronouns and verbs in ASL provides mapping instructions for their semantic structures. This leads to a type of meaning construction in which the mental spaces that serve as the basis for meaning construction are semantic structures. If such mappings are generalized to other instances of language use, both spoken and signed, it appears that the benefits of mental space mappings are maintained. Semantic structures are an indispensable component of the chain linking phonetics to comprehension. The data discussed in this paper suggest that meaning can be satisfactorily constructed if built around these obligatory representations.

## References

- Edge, VickiLee, and Leora Herrmann. 1977. Verbs and the Determination of Subject in American Sign Language. In L. Friedman (ed.), *On the Other Hand*. New York: Academic Press, 137-179.
- Fauconnier, Gilles. 1985. *Mental Spaces*. Cambridge, MA: MIT Press. Reprinted 1994, Cambridge, UK: Cambridge University Press.
- Fauconnier, Gilles. 1997. *Mappings in Thought and Language*. Cambridge, UK: Cambridge University Press.
- Fauconnier, Gilles, and Mark Turner. 1996. Blending as a Central Process of Grammar. In Adele Goldberg (ed.), *Conceptual Structure, Discourse and Language*. Stanford, CA: CSLI Publications, 113-130.
- Fischer, Susan. 1975. Influences on Word Order Change in ASL. In C. N. Li (ed.), *Word Order and Word Order Change*. Austin, TX: University of Texas Press, 1-25.
- Friedman, Lynn. 1975. Space, Time, and Person Reference in American Sign Language. *Language* 51:940-961.
- Gee, James, and Judy Kegl. 1982. Semantic Perspicuity and the Locative Hypothesis: Implications for Acquisition. *Journal of Education* 164(2).

*Grammar and Gesture in American Sign Language*

- Klima, Edward S., and Ursula Bellugi. 1979. *The Signs of Language*. With R. Battison, P. Boyes Braem, S. Fischer, N. Frishberg, H. Lane, E. M. Lentz, D. Newkirk, E. Newport, C. C. Pedersen, and P. Siple. Cambridge, MA: Harvard University Press.
- Lacy, Richard. 1974. Putting Some of the Syntax back into Semantics. Ms., The Salk Institute.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, Vol. II: Descriptive Application*. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 1999. Discourse in Cognitive Grammar. Paper presented at the International Cognitive Linguistics Association. Stockholm, Sweden, July 1999.
- Liddell, Scott K. 1980. *American Sign Language Syntax*. The Hague: Mouton.
- Liddell, Scott K. 1994. Tokens and Surrogates. In Inger Ahlgren, Brita Bergman, and Mary Brennan (eds.), *Perspectives on Sign Language Structure: Papers from the Fifth International Symposium on Sign Language Research, Vol. I*. University of Durham, England: The International Sign Linguistics Association, 105-119.
- Liddell, Scott K. 1995. Real, Surrogate, and Token Space: Grammatical Consequences in ASL. In Karen Emmorey and Judy Reilly (eds.), *Language, Gesture, and Space*. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, 19-41.
- Neidle, Carol, Dawn MacLaughlin, Judy Kegl, Ben Bahan, and Debra Aarons. 1995. Overt Realization of Syntactic Features in American Sign Language. Syntax seminar, University of Trondheim, Norway, May 1995.
- Padden, Carol. 1988. *Interaction of Morphology and Syntax in American Sign Language*. New York: Garland Publishing.
- Stokoe, William C. 1960. Sign Language Structure: An Outline of the Visual Communication System of the American Deaf. *Studies in Linguistics Occasional Papers, No. 8*.
- Woodward, James. 1970. Personal Pronominalization in American Sign Language. Ms., Georgetown University.

Department of Linguistics  
Gallaudet University  
Dawes House  
800 Florida Avenue, NE  
Washington, DC 20002

scott.liddell@gallaudet.edu



# Prosodic and Gestural Cues for Navigations around Mental Space

JUNE LUCHJENBROERS

*University of Wales at Bangor*

## 0. Introduction

This paper reports on research that explores how lexical, prosodic, and gestural information combine to provide discourse participants with the appropriate cues needed to set up and structure mental spaces.<sup>1</sup> The particular issues considered here include: (i) the role of prosodic stress in alerting hearers to the necessary movements around conceptual space, (ii) the role of gestural information in navigating conceptual space, and (iii) the extent to which prosodic and gestural cues converge to help the hearer to navigate the speaker's use of mental space. This paper builds on earlier work that has used information flow categories as a predictor of attentional focus in mental space construction and navigation. This research agenda now includes audio and video data of negotiated talk, which has raised important, new questions about the primacy of information flow categories as predictors of mental space activity and about the facilitation of the discourse building process by prosodic and/or gestural information.

## 1. The Data

### 1.1. Research Design

The data used for this paper comes from a larger videotaped study into negotiated talk between one Australian male or female student, discussing behaviors to be avoided at university, and either one other Australian or foreign student. A total of 32 interactional dyads was collected.

Subjects were recorded in a soundproof room positioned diagonally opposite each other. The purpose of this positioning was to maximize the view for the analyst (sitting in the next room behind a large tinted window) and the video recorder without drawing undue attention to either. Subjects reported that they found the analyst easy to ignore, but that was less true of the video recorder.

---

<sup>1</sup> The research drawn upon in this paper was supported by a postdoctoral grant (supplemented by grant #122 3360 04 000 R CLTR/UQNSRG088G 97) to the author with the University of Queensland (Australia). I thank Eve Sweetser, Satoro Kita, and Susan Duncan for their feedback at the presentation, and Wilbert Kraan for running his eye over the draft manuscript. All errors are mine.

The participants in each dyad were given the task of devising guidelines (to be given to faculty) about how new students should avoid the pitfalls associated with (a) cheating, or (b) plagiarism. Each dyad lasted roughly 30 minutes.

Subjects were drawn from any discipline and at differing levels of study (i.e. both undergraduate and postgraduate students were included). Participants had either not previously met or did not know each other well, and they were paid for their participation.

### **1.2. Initial Observations**

This paper focuses on 12 of the 32 interactional dyads, which is the total number of dyads involving only Australian students. The Australians-only data is made up of 4 male-male dyads, 4 male-female dyads, and 4 female-female dyads.

The first discovery made in these data is that there are remarkable individual and gender differences in how much subjects make use of gesture during conversation. Australian women gesture far more than Australian men, and noticeably more so when talking to another Australian woman than when talking to an Australian man. Australian men gesture very little, and particularly so when talking to another Australian man.

I have therefore found it more practical to first find good gesture examples and establish how lexical and prosodic cues fit in, and then to start at the beginning of a conversation and see how the chosen factors converge. In general, the examples given in this paper were drawn from female-female dyads. In contrast, analysis of prosodic information can easily start at the beginning of talk because lexical and prosodic information go hand in hand.

## **2. Prosodic Stress and *Focus***

The chosen prosodic feature for this project is stress assignment, and the research objective is to determine how stress conveys mental space information. Like others in the discourse field, I view discourse as a process of mutual ground construction where speakers are responsible for giving addressees adequate cues to derive their intended meanings, and addressees are responsible for making a determined search for those meanings. This discourse building process has been largely mapped using the range of functionalist tools at the analyst's disposal: *focus*, *topic*, *newness*, and *givenness* (cf. Chafe 1994, Lambrecht 1994, Luchjenbroers 1993). However, notable difficulties exist with the definitions of these tools and with adequately defining the relationships between them (Luchjenbroers in press, in progress).

In particular, prosodic stress most closely associates with the notion of focus; however, focus, like many other concepts in the discourse analyst's armory, is vague and often misused. Discourse elements may achieve focus in the mind of the hearer, and those same elements may be stressed in the speaker's attempt to give hearers adequate cues about speaker-intended meanings, but the two processes are not the same. In essence, what happens in the speaker's mind ought not to be confused with what is thought to happen in the hearer's mind. In my work,

stress assignment is a hearer-oriented, speaker strategy, and focus is a property of the hearer's cognitive system.

The second major association for stress is with the information flow category of new/newsworthy information. I have found this to be a largely reliable association in that new/newsworthy information tends to attract increased stress in the speaker's output; however, information flow, like focus, is also a property of the hearer's cognitive system.

In effect, speakers have control over lexical choice, grammatical placements, and prosody, but the key conceptual processes with which these linguistic strategies are generally associated (focus and information flow) belong to the hearer's cognitive system. The communication process relies on speakers using those cues within their control (e.g. stress and lexical placement) to enable the hearer to follow the speaker-intended meaning. This involves speakers making assumptions about the likely information flow status that information would have in the mind of the hearer in order for them to adequately structure their outputs—hence, the discourse analysts' tendency to collapse speaker stimuli with expected hearer cognitive phenomena. However, even preliminary analyses reveal that much more information receives focal treatment (i.e. is stressed) than is new or newsworthy information. For further clarification of these processes, this work looks to Mental Spaces theory (Fauconnier 1985/1994).

### **2.1. Mental Space Navigations**

Within the Mental Spaces approach, contributions consist of two fundamental elements: spaces and those propositions to be processed within them. Examples of how these two elements are unambiguously separated can easily be found. For example, in (1), truth of the queried proposition *is that an acceptable thing* is clearly limited to the locative space *in Sabah, in Borneo*. If the speaker were to have used a different spatial definition (e.g. *in Queensland*), she would have changed the context in which the proposition can be measured as true or false.

(1) H: *is that an ac'ceptable **thing** ↑ in in in **Sabah** ↑ in **Borneo** ↑<sup>2</sup> [15-16]*

This example illustrates that primary stress (shown as bold text) occurs in both in the propositional component and the spatial definition.

In contrast with (1), however, there are cases where the distinction between proposition and space is less obvious. For example, in (2) a number of locative phrases provide additional information to the propositional component, but do not function as the contextual spaces in which the propositions are to be processed. The breakdown of Proposition to Spaces for examples (1) and (2) is given in (3).

---

<sup>2</sup> Examples are given with the following additional information: bold = primary stress (prosodic pulse); underlined = where verbal and gesture components coincide; arrow ↓ above text = gesture onset; arrows (↑ ↓ → within text examples) = intonation contour. Note that an arrow → in front of an example indicates the particular line to which attention is being drawn.

- (2) E: *I had some.. ah overseas students staying with me* → [8]  
*they'd just come from. Sabah* ↑ *in Borneo* ↑ *and um* [9]  
*they were in foundation year* ↑ [10]

| (3)       | <b>Propositional Component + Pointer</b> | vs. | <b>Space Builder</b>   |
|-----------|------------------------------------------|-----|------------------------|
| [line 8]  | <i>I had overseas students staying</i>   |     | <i>with me</i>         |
| [line 9]  | <i>They'd just come</i>                  |     | <i>from Sabah</i>      |
|           |                                          |     | <i>in Borneo</i>       |
| [line 10] | <i>They were in</i>                      |     | <i>foundation year</i> |
| [line 15] | <i>Is that an acceptable thing</i>       |     | <i>in in in Sabah</i>  |
| [line 16] |                                          |     | <i>in Borneo</i>       |

Examples such as these illustrate that although locative phrases are prime candidates for space builders, they do not always function as such; however, I would argue that their spatial nature is not entirely lost, and in a sense act as “pointers” to potential spaces outside the immediate context of talk (Luchjenbroers 1993). At this point I offer the suggestion that the first mentions in lines 8-10 are references to locations “outside” focus (“F-space”), while the second reference in lines 15-16 becomes “inside” focus, as it here functions as the contextual space in which the new proposition is to be considered. The difference between pointers and spaces is that the latter add structure to the developing discourse network (cf. Sweetser and Fauconnier 1996).

Of interest also is that these locative phrases attract prosodic stress in both examples (1) and (2). This gives evidence that the presumed correlation between new information and stress is not two-way: although what the speaker predicts to be new information to the hearer attracts stress, stress clearly must correlate with more than new/newsworthy information. I offer the suggestion that the repeated stress in the later talk (later by only five main clauses) serves to amplify a shift in function from the propositional component (line 9) to the space in which the new proposition is to be processed (lines 15-16). Hence, prosodic stress signposts that information either (i) is new, whether spatial, pointer, or propositional, or (ii) has changed function, such as movement from proposition to spatial definition. Alternatively, following the suggested terminology above, prosodic stress also signposts that information has moved from outside to inside F-space. Preliminary observations suggest that movement in the other direction (from inside to outside F-space) does not attract prosodic stress; in such cases the general rule of “given information is not focal” is seen to apply.

A final observation is that when an active space is altered in some way, the altered part will likely attract stress, e.g. (4).

- (4) E: *they 'didn't have much experience in essay writing in* → [20]  
*(in breath) in the English style* ↑ [21]

In (4) the spatial reference *in essay writing* (line 20) is a first mention and is in focus but does not attract stress; presumably because it is so deeply associated

with the primary topic (plagiarism) it is considered active. In contrast, the subsequent spatial modification *in the English style*, which should otherwise be unstressed for the same reason, contrasts with the spatial context of talk before line 20, *practices in Borneo*, and is therefore stressed. There is good reason to utilize the known phenomenon of ‘contrastive stress’ for mental space maneuvers, and this will emerge in later discussions also, but arguably only when information moves into focus.

## 2.2. Discourse ‘Levels’

A full consideration of the data has also prompted a consideration of the embeddedness of the spaces referred to in talk. In particular, at least four distinct levels (or layers) have become apparent (see also Rubba 1996). These four levels, illustrated in Figure 1, are defined as follows.

Level 1: the **here and now** context of talk, which is the dynamic of being in an enclosed area, talking to a fellow student who they do not know well (if at all), in front of a video camera and an academic who is taping every word. This is the most primary space (cf. Base Space), and talk at any time can revert to this layer, e.g. line 349 in (5).

- (5) → F: *gosh I sure like this pen* ↓ [349]  
S: *I like it too (both giggle).... umm..* [350]  
F: *ok.. what about exams now.....* [351]

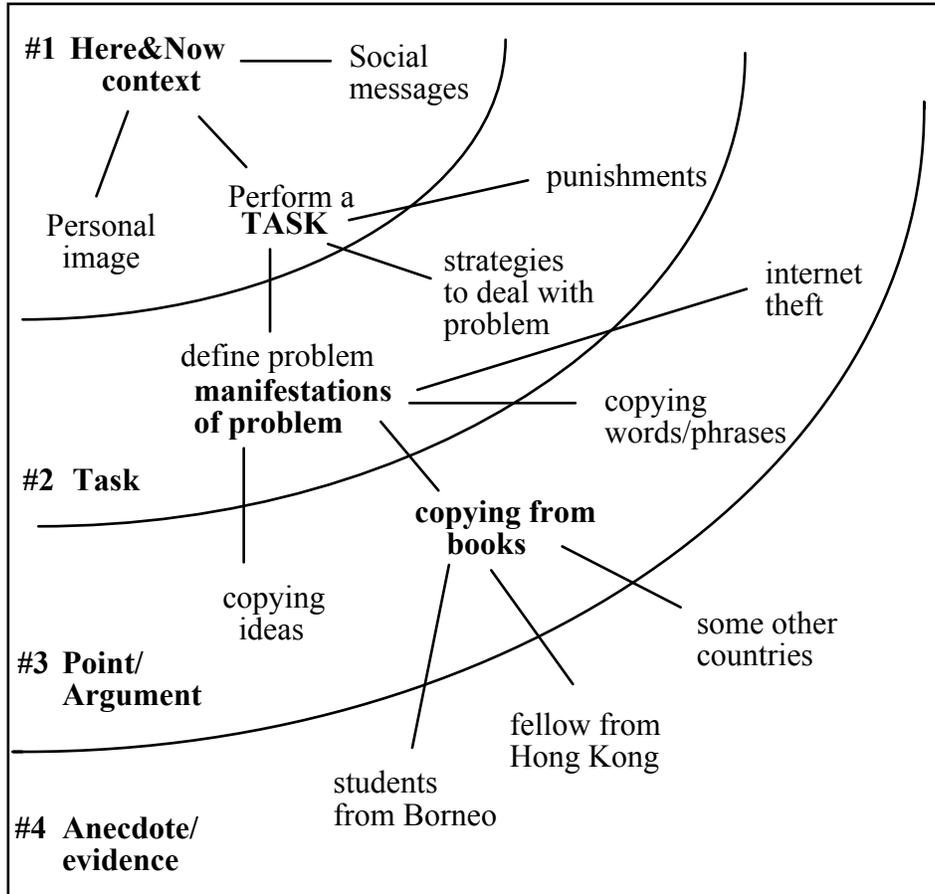
Within this general context layer, speakers need to contend with a number of factors in addition to the task being performed. For example, they need to deal with the interpersonal dynamics of the situation, such as social rules of interaction (how to proceed), as well as live up to personal ideals, such as ‘look intelligent’, and ‘don’t let the other person get the better of you’, etc. Notably, line 351 takes talk back to the next level of talk, Level 2.

Level 2: refers to the Task speakers are involved in. At this level, speakers negotiate how they will proceed, e.g. (6). It is a level of interaction that carves up the subject matter into manageable chunks.

- (6) H: *I guess I guess. what we **should** do is is um..* ↓ [5]  
→ ***de’c-ide** what **plagiarism** is first* → [6]

Dealing with the task involves a range of issues, including defining the problem (which in these data takes the form of recognizing how the problem is manifested), producing strategies for dealing with the problem, and identifying the consequences of being found guilty. The abstract examples of these Level 2 categories make up the next level of talk, Level 3.

**Figure 1.** Embedding of Levels



Level 3: includes the **Points** or **Arguments** that emerge from dealing with the Task, e.g. (7). These are general statements that define or explain the tasks decided upon in the Task layer.

(7) E: *oh it's.. more or less whole sale copying from books* → [8]

Level 4: is the final level and includes **Anecdotal evidence** of the Points or Arguments given in talk, e.g. (8).

(8) → E: *I had some.. ah overseas students staying with me* → [9]  
*and they used to copy. whole sale..pieces from books* → [12]

Here again, specific evidence defines or explains individual (abstract) Points (Level 3) which combined make up different aspects of the Task (Level 2), and everything to do with the task is a part of what the speaker is dealing with in the here-and-now of talk (Level 1). It may be difficult to imagine that speakers keep track of where they are in the tree at all times; nevertheless, if a hearer were to

*Cues for Navigations around Mental Space*

confuse how a piece of evidence proves a Point, or how a Point was relevant to the Task being performed, either miscommunication would follow or comprehension would fail.

Possible evidence for this Levels approach may be provided by accompanying stress. In particular, a Levels analysis may clarify why some (otherwise given) spaces attract stress when a simple approach to either spaces or information flow might predict otherwise, e.g. (9).

- (9) E: *oh it's.. more or less whole. sale copying from books* → [8]  
.....  
H: *so i-i-if we sa-ay. first of all that plagiarism* → [26]  
→ *'s just whole sale copying of stuff from books an that* → [27]

In (9), the lexical element of interest is the propositional component (*Pointer from books*) which is stressed in these two mentions (lines 8 and 27, said roughly 35 seconds apart). The information is not new and in such a short time span is certainly still active; however, movement around the network is complex. The contribution in line 8 functions from the Point Level, and between lines 8 and 27 talk moves from that Level to Evidence #1, up to the Task Level and then to Evidence #1 again, followed by Evidence #2, and then in line 26 back to the Point Level. Prosodic stress continues where the lexical component is given, but the mental space in which it must function changes. I suggest that continued prosodic stress of such active components can be explained as a speaker cue to hearers about how to navigate conceptual and discourse space.

The Levels approach suggested here also complements the inside vs. outside F-space suggestion offered above. In line 27 the element (*copying from books*) is moved back into focus. Hence this example illustrates how prosodic stress, Levels, and inside focus converge.

Observations of the data so far have led to the following generalizations:

- a. New or newsworthy Information (whether as Space or Proposition) will likely attract primary stress.
- b. When Propositional information becomes the spatial definition of a subsequent Proposition (i.e. moves from outside to inside F-space), it will likely attract primary stress.
- c. When spatial information is modified in some way (i.e. moves into F-space), the modified part will likely attract primary stress.
- d. When information (whether Space or Proposition) signposts a different Level in talk (i.e. a different Level moves into F-space), it will likely attract primary stress.

In these examples I have shown that stress assignment goes beyond new information and is an important cue to hearers about how to maneuver very intricate

aspects of discourse structure. In the following section I will consider how gesture may be said to join in.

### 3. Gestural Cues

When speakers gesture, the gestures they produce take place within a comfortable physical space in front of them. The dimension of this space is roughly the shape of a cube that runs from shoulder to waist in height, from elbow (at the waist or, in these data, the table) to the hand in depth, and has body width. The actual size of a speaker's gesture space (cube) varies from speaker to speaker, and very likely from culture to culture, as does the proportion of gesture to speech. In general, speakers who are less animated in gesture use a smaller gestural cube, and those who are more animated use a larger cube that is more consistent with the dimensions given. The size of a speaker's gesture space is defined by where they make most of their gestures. However, in addition to where speakers make most of their gestures they also make numerous gestures that are clearly outside these general boundaries. I suggest that these general vs. extreme boundaries are consistent with inside vs. outside gestural F-space, and when a gesture is made within the cube, or not, the speaker is conveying additional but relevant information about navigating mental space.

To further clarify this suggestion, gestures are considered from a number of angles. In particular, this research has considered gesture in terms of (i) gestural types, (ii) gestural complexity, and (iii) how types and complexity convey information to the hearer about mental space navigations.

#### 3.1. Gestural Types

Researchers in gesture (see McNeill 2000) generally recognize at least three types of gesture: Deictic or Indexical gestures, Iconic gestures, and Pantomimes. In this paper I will discuss Indexical and Iconic gestures. Indexicals are the most basic form of gesture and refer specifically to gestures involving (a) only the index finger, e.g. (10), (b) the full hand, e.g. (11), or (c) an instrument such as a pen, e.g. (12).

- (10) E: *there's even a special **section** of legal studies at QUT* [178]  
 L arm crosses the body (& F-space)  
 L finger points away from S in direction of QUT ( $\neq$  F-space)
- (11) H: *but... of course at the university that's. that's **not on**  $\uparrow$*  [41]  
 both hands in a flat cup, palms down & slanting inwards,  
 fingers touching the table, in front of S (= F-space)
- (12) E: *they say the university policy here .. is* [57]  
 pen points down, touching the table, in front of S (= F-space)

In all three examples the physical location of the referent has a direct relation to the physical location pointed to by the speaker. Also, in terms of F-space, the

### *Cues for Navigations around Mental Space*

relation between the physical locations of the referents and the indexical gestures is no coincidence. The gestures given in examples (11) and (12), referring to “here”, are firmly in the center of the speaker’s gestural space (i.e. in F-space), but example (10), referring to a different university, is clearly outside it. This gestural choice is not arbitrary; the speaker could have chosen the right hand to make this gesture to the right (when writing the speaker is also right-handed), in which case the indexical would still have been outside F-space but that choice might have been coincidental and therefore ambiguous. In this case the speaker’s choice to use the left hand to cross the body (and F-space) to a position that is again outside the speaker’s gestural F-space is more telling of the speaker’s intent and the focal status of that information. The body of talk is about practices at university (in Australia), but the gestures make clear that for these speakers, the specific space that is maximally active (and focal) is what happens “here” (at the University of Queensland) as opposed to “not here”.

Notably, all three examples are also examples of contrastive stress. In section 2 of this paper, stress was associated with bringing elements into focus, either because they convey new information or because they involve processing at a different conceptual level from the contribution before. Examples (10)-(12) reveal a similar process in that in these examples stress amplifies a change in spatial definition to spaces that are either new or renewed. However, these (re-) activated conceptual spaces also deal with locations in physical space, and the gesture placements either inside or outside F-space amplify the relation of those locations to the speaker. Hence, Indexical gestures go beyond lexical instructions for mental space creation and identification.

Similarly, Iconic gestures can function like Indexicals (i.e. point to physical locations), although the relation between the gesture and the referent is not always straightforward. Iconic gestures often convey information that is relevant to more than one domain in mental space navigations, e.g. (13a).

(13) a. H: *that’s... completely **peppered** with.. you know t t t t t* [249]  
R hand, pinched (all four fingers on top of thumb), pointing  
to a number of points very close together, above R eye ( $\neq$  F-space)

b. In this hypothetical student paper, there are many plagiarized bits.

In (13a) the speaker points to locations in physical space (above F-space). The Iconic gestures represent the unseen pages of a typical student paper that has been plagiarized, and the points in space refer to locations in that hypothetical paper. The frequency of the jabbing motions to locations in the air correlates with the frequency of plagiarized sections in that student paper. The gestures are clearly locational but the information is part of the propositional component within the space that is created by virtue of the gesture. An extended interpretation of this sequence would be a verbalized string such as given in (13b). The space builder *in this hypothetical student paper* is created to make sense of the gesture, and the proposition *there are many plagiarized bits* is conveyed by the high number of

jabs in space, and serves as an elaboration of the proposition already verbally conveyed. Hence the gesture conveys both spatial and propositional information.

Another important feature of Iconic gestures that appear to have an Indexical character is that they may be relevant to recognizing the Levels (of embeddedness) in talk referred to earlier, e.g. (14) and (15).

- ↓
- (14) H: *so i-i-if we **sa-ay**. first of all that plagiarism →* [26]  
 → *'s just whole sale **copying of stuff from books** an that →* [27]  
 L hand serves as a list, R hand points to fingers, each referring  
 to a different Point for the Task (define the problem) (= F-space)

- ↓
- (15) H: *even even copying someone else's assignment* [105]  
 index finger pointing at H (= F-space)

Examples like (14) and (15) involve pointing gestures that do not refer to the content at all. These are interactional because they serve as a kind of instruction to the hearer. In (14), the gesture instructs the hearer that the speaker's contribution is a Point to be added to the list of Points. Similarly, in (15), the pointing finger is not an accusation that the hearer is guilty of the said act, but here the gesture also functions as an instruction to the hearer to add a new Point to the list of Points she is jotting down. In cases such as these, gestures instruct hearers to shift Levels; they are not for clarifying or complementing content.

It is clear from examples like (13)-(15) that gestures can convey different kinds of information to the lexical component. Similarly, in (16) it is difficult to categorize the associated gesture as spatial or propositional.

- ↓                    ↓                    ↓
- (16) H: *so we 'say if you do.. **this.. and this and this*** [28]  
 R hand, pinched (all four fingers on top of thumb), pointing to  
 a number of equidistant points in space, forming an oblique row  
 from just above F-space (height of L eye), into F-space (below R  
 shoulder)

In (16) the speaker uses pointing gestures to possible events (that constitute plagiarism) as locations in physical space. In this example the relationship between gesture icon and referent is more abstract than any of the earlier examples as the gestures refer to possible events and not locations. In examples (14) and (15) the location is arguably either the list that the speaker has made with her fingers, or the list of Points (Arguments) the other person is writing down. However, the pointing gestures in (16) refer to Points made earlier in talk, and which can reasonably be expected to exist in conceptual structure. These pointing gestures therefore are locational but in a more abstract sense than the earlier examples.

The complexity of this example is greater than deciding whether to analyze the gesture as spatial, pointer, or proposition, as it also refers the speaker back to another Level of talk: the Task Level (having just discussed more evidence). The Points/Arguments in conceptual space, indicated by *this.. and this and this*, were all produced with prosodic stress, which would be expected when Levels jump back into focus, as well as when contrasting discourse elements are specified. Nevertheless, the conceptual load in processing all informational cues is enormous, the complexity of which is missed by just an analysis of the lexical component.

### 3.2. Gestural Complexity

As already evident in the examples above, gestures can very complex. The data also contains many ‘simple’ gestures (i.e. those that convey a straightforward semantic relationship between the essential message carried by the gesture and the lexical component it accompanies). For example, a frequent simple gesture used during these talks is a *take* gesture (one hand scoops an unseen substance or object and draws it to the body), which was used to complement talk of *taking*, *stealing*, *plagiarizing*, and *cheating*. Such gestures are simple because they are consistent with the verbal component. However, examples where gestures point to iconic locations, such as in (13) above, or events, such as in (16) above, are complex because (among other things) the physical location of the gestures holds no straightforward relation to the referents. In the latter cases, gestures do substantially more than clarify lexical meanings; they complement the lexical component by providing meanings the speaker does not (need to) articulate.

Complex gestures also vary in the range of complexity involved, from the very complex examples already given to less complex examples where the gesture meaning is closely related to the meaning conveyed lexically, e.g. (17).

- (17) H: *like-t.. it’s like, it’s some sort of public.. **public information*** [38]  
both hands splayed, palms facing each other, making circles  
opposite each other (= F-space)

In (17) the gesture associated with *public* suggests “mixing”, which conveys an attribute that the word does not entirely convey. Here the sense that is added is that public information is a mixture from multiple sources (who cares where?), and it is therefore legal to mix that information with one’s own resources. Example (18) below also utilizes the “mixing” gesture, but is more complex because it also conveys additional information through the use of F-space.

- ↓
- (18) E: *so.. he just couldn’t.. **turn that information around..*** [264]  
both hands, palms facing S, rotating from away from S,  
up and over the other hand to closer to S, several times  
(= mixing outside F-space into F-space)

Example (18) involves a component of meaning that is not only clarifying of the lexical component, such as in (17), but is also complementary. The alternating positioning of the hands correlates with the lexical component *turn around* but the gesture conveys scooping an invisible substance toward the speaker in a large circular motion and involves moving the hands from outside F-space into F-space. The directionality of this gesture is significant because the resting state for each hand is roughly the center of the chest, which clearly conveys “mine” (or, in this case, “make mine”). This is a clear example of how the boundaries of F-space can be used to complement both the semantics of the lexical component as well as the gesture. This sense extension relies on the contrast between inside F-space = “here” and “me/mine” vs. outside F-space = “not here” and “not me/mine”.

In sum, gestural complexity involves multiple types of information: spatial, propositional, and sometimes interactional. Unlike the lexical component which can generally be unambiguously assigned one or other mental space role (i.e. space builder or proposition), gestures often contain components with multiple roles. The issue remaining is to determine how gestural and prosodic features may be said to combine to disambiguate the complexity of the hearer’s input.

### 3.3. Prosody and Gesture: Mental Spaces Cues

Mental spaces require spatial definitions in which propositions can be processed. However, the kinds of information both prosodic and gestural information provide has already been shown to advance on these two elements, with some clear correlates with Levels of embeddedness, as well as gestures that appear to convey both spatial and propositional information. In addition, gestures sometimes provide interactional information to the hearer. In examples (14) and (15) above, these were instructions relating to contextual Levels, but other examples reveal how gestures combine with prosodic information to achieve the same interactional outcome, e.g. (19), where the gesture together with an increase in stress and slowing of speech indicates that the speaker has completed her turn and is offering up the floor.

- (19) E: *n also there’s a fellow there from **Hong Kong** you know...* [23]  
 → *he was reel-ly ba-ad.* [24]  
 nodding head (= end of turn)

In contrast, however, other examples have revealed how prosodic stress and gestural information appear to convey very different kinds of information. In the Indexical gesture examples (10)-(12), prosodic stress amplified a change in spatial definition while the gestural placements, either inside or outside F-space, amplified how those locations relate to the speaker. Essentially the gestural and prosodic cues in these examples provided very different kinds of information that serve to offer greater depth to the information offered in the speech stream.

The range of examples presented provide a number of meanings for F-space. In terms of prosodic stress, focus is related to speaker assumptions about what would be new information to the hearer, and to changes in the location in which

### *Cues for Navigations around Mental Space*

an utterance operates within the mental spaces network, which in some cases is represented as Levels of embeddedness. In terms of gesture space, however, focus refers to either the micro-context in which a proposition is to be processed (akin to space builders or the propositions within them), or to another important form of contrast, that between aspects more closely associated with the individual (“here”, “me”, and “mine”) vs. aspects that are not (“not here”, “not me”, and “not mine”).

In addition to this egocentric nature of a person’s gesture space, however, gestures often convey important information about spatial definitions and also often amplify changes in them, e.g. (20).

- (20) H: *I 'think... I 'think sn 'some other countries* [34]  
hands together (prayer posture) in front of S (= F-space)
- *well not **just** other countries* [35]  
*but I I 'think some 'people um..* [36]  
hands move apart from prayer posture, in front of S (= F-space)

After a space has been created, an alteration of that space will involve a change in hand gesture. In (20), line 34 is said with both hands together, which does not indicate anything about the space (other than maybe illustrating the speaker’s contemplation), but the change in space in line 35 motivates the speaker to move her hands apart. This is not a coincidence; the space has been broadened (widened) to include more people and the space between the hands has likewise been increased. The hearer is also alerted by the stress on *just* which coincides with the change in gesture position.

#### **4. Is This a Concert?**

In this paper I have presented a number of examples to suggest that the answer to this question is a tentative “yes”. I make this suggestion tentatively because although there are many cases where the informational load of gestural information substantially enriches the speech stream to which a hearer is exposed, there are also numerous cases where gestural information substantially precedes the lexical component, in which case there is reason to suppose that gestures might often be indicative of speakers making their own way through mental space, as they are hearer-oriented cues to facilitate comprehension during discourse.

There is also clear evidence of a division of labor between prosodic and gestural information, which often converge and complement each other by enriching the information potentially available to the hearer. Stress assignment is clearly related to presumed accessibility of the spatial component of an utterance and therefore plays a role in informing hearers about navigating space in terms of how far they need to search (or how far the speaker wants/expects them to search). In terms of gesture, however, the boundaries between space and proposition are sometimes blurred as multiple types of information can simultaneously be conveyed. In particular, gestures may clarify an element that belongs within the propositional element, but may be produced in a physical space that lies outside gestural F-space; therefore, they might convey spatial information about the

location. Similarly, in some cases, gestures convey information about what level that information pertains to (e.g. examples or task). In sum, the complexity of gestures is scarcely understood, and much more data needs to be analyzed to alter that state. It is, however, clear from the examples presented in this paper that the role of gestures (together with prosodic information) is a fruitful and important area of investigation for a Mental Spaces approach to discourse.

## References

- Chafe, Wallace. 1994. *Discourse, Consciousness, and Time*. Chicago: University of Chicago Press.
- Fauconnier, Gilles. 1985/1994. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. New York: Cambridge University Press.
- Lambrecht, Knut. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Luchjenbroers, June. 1993. Pragmatic Inference in Language Processing. Ph.D. diss., La Trobe University, Melbourne, Australia.
- Luchjenbroers, June. In press. "Functionalism, Conceptual Mappings and Mutual Ground Construction in Court." In P. Casanovas (ed.).
- Luchjenbroers, June. In progress. "Conceptual Mappings and Mutual Ground Construction." In A. Verhagen and J. v.d. Weider (eds.), *Levels in Cognition*.
- McNeill, David, ed. 2000. *Language and Gesture*. Language, Culture, and Cognition. Cambridge: Cambridge University Press.
- Rubba, J. 1996. "Interpretation of Deictic Expressions." In G. Fauconnier and E. Sweetser (eds.), *Spaces, Worlds and Grammar*. Chicago: University of Chicago Press.
- Sweetser, E. and G. Fauconnier. 1996. "Cognitive Links and Domains: Basic Aspects of Mental Spaces Theory." In G. Fauconnier and E. Sweetser (eds.), *Spaces, Worlds and Grammar*. Chicago: University of Chicago Press.

Department of Linguistics  
University of Wales at Bangor  
Gwynedd, LL57 2DG, Wales  
United Kingdom

junel@bangor.ac.uk

# **Does Gesture Help Processes of Speech Production? Evidence for Conceptual Level Facilitation**

ALISSA MELINGER<sup>1</sup> and SOTARO KITA<sup>2</sup>  
*Saarland University<sup>1</sup> and Max Planck Institute for Psycholinguistics<sup>2</sup>*

## **0. Introduction**

When people speak, they gesture, and these gestures are often transparently related to the semantic content expressed in the speech. Gestures that are semantically co-expressive with the co-temporal speech are called representational gestures. Representational gestures can encode various semantic features of objects and events, such as the shape of an object, the interactive characteristic of an object, the function of an object, an activity, an entity's spatial location, etc. While many have claimed that representational gestures serve a communicative function (see Kendon 1994 for a survey), it has also been observed that gestures are produced even in the absence of a visible interlocutor. When people speak on the phone or to a non-visible companion, they continue to produce co-speech gestures (Cohen 1977). This suggests that these gestures may serve a(n additional) purpose that is not communicative. Some researchers have proposed that gestures might actually aid processes of speech production; however, there is a great deal of disagreement about how this might work (see Kita 2000 for a review of the debate). This paper investigates the issue of how gestures aid speech production processes.

## **1. Stages of Speech Production**

It is generally agreed that processes of speech production are incremental and can be broken down into three processing stages: (a) conceptualization, (b) formulation, and (c) articulation (Levelt 1989). According to Levelt, at the conceptualization stage the speaker conceives of an intention, selects the relevant information to be expressed, and orders the information for expression, among other processes. It is within the conceptual level that “thinking for speaking”, as intended by Slobin (1987), occurs. At the formulation stage the pre-linguistic message is given linguistic form. Lexical items and syntactic frames are selected, ordered, and combined at this stage. Finally, at the articulation stage signals are transmitted to the articulators to produce the desired utterance.

Additionally, there is a large body of evidence from several distinct sources suggesting that the formulation stage is further divided into two distinct levels of

representation in which different types of information are retrieved. First, an abstract lexico-semantic/syntactic representation (or lemma, cf. Kempen and Huijbers 1983) is retrieved, based on the semantic information passed down from the conceptualizer. In a separate stage of processing, the form of the word is specified. At this word form level, the metrical, segmental, and morphological structure of the word is specified.

## **2. What Role Does Gesture Play for the Speaker?**

Many researchers have speculated on how gestures might contribute to the general processes involved in speech production. Many researchers propose that gesture facilitates speaking by aiding in the process of lexical retrieval, i.e. during formulation. We will refer to this hypothesis as the Lexical Level Hypothesis. Specifically, the claim is that a gesture serves as a cross-modal prime to boost the activation of a particular lexical entry, either at the lemma level (Krauss et al. 1996) or the word form level (Butterworth and Hadar 1989, Krauss et al. 2000). They cite the fact that when gesture production is restricted, speech including spatial content is adversely affected (Rauscher et al. 1996). Furthermore, gesture prohibition has also been shown to increase the number of retrieval failures in a tip-of-the-tongue elicitation study (Frick-Horbury and Guttentag 1998). Proponents of the Lexical Level Hypothesis interpret these results as evidence that gesturing facilitates lexical retrieval.

Another interpretation of these same data, however, is that gesture prohibition adversely affects the conceptual processes involved in constructing a pre-linguistic message rather than in lexical retrieval processes per se. If gestures aid in the activation of conceptual representations, this activation will then spread to the lexical level. Thus, effects such as those used to support the Lexical Level Hypothesis are also consistent with a model in which gestures aid processes at the conceptual level. Supporters of this view argue that gestures help activate imagistic and conceptual information at a pre-lexical level (de Ruiter 1998) and help to map between imagistic information and propositional information (Alibali et al. 2000). We will refer to the hypothesis that gesture facilitates conceptual level processes as the Conceptual Level Hypothesis.

One problem with both the lexical and conceptual level hypotheses is that certain details crucial to formulating testable predictions are underspecified. For example, very little is known about the general conceptual processes that are involved with speaking, and therefore it is difficult to be explicit about how gesture could facilitate these processes. On the other hand, it is also difficult to understand exactly how a gesture could prime lexical information at, for example, the word form level, given that evidence from speech production literature suggests that semantic information is not represented at that level. In other words, why should making a circular motion with a finger help the retrieval of a representation that specifies the number of syllables or the segmental content for the word circumference? It seems more likely that a gesture should prime semantic representations rather than phonological representations. However, even claiming that gestures

prime semantic representations is compatible with both hypotheses because there is a great deal of debate in the production literature as to whether lemmas contain any semantic information themselves (Butterworth 1989) or whether they only have meaning by virtue of connections to the conceptual level (Levelt et al. 1999, Dell 1986). Furthermore, given that lemmas are activated as the result of activation passed on from the conceptual level, it is quite difficult to distinguish effects that occur as a direct link between gesture and the lexicon from effects that have their locus at the conceptual level and then trickle down to the lexical level. In fact, these two possibilities are so indistinguishable that most researchers who propose the latter must also allow for the possibility of the former (Krauss et al. 1996, Krauss et al. 2000).

Despite the difficulties in distinguishing these two hypotheses, they do predict a different distribution of gesture-speech interactions; they predict that linguistic differences will have an effect on gesture under very different circumstances. For example, the Lexical Level Hypothesis predicts that gestures are produced when lexical retrieval is more difficult due to inherent characteristics of the target word, such as having low frequency or many lexical competitors, or when retrieval is more difficult due to external or contextual factors. For example, if a word is produced twice in immediately adjacent clauses, the second mention should be easier to retrieve than the first mention, since the lemma should have some residual activation from when the word was first produced. The relative ease of producing the second mention should result in fewer co-expressive gestures than were produced for the first mention if gestures are produced to facilitate lexical retrieval. In contrast, the Conceptual Level Hypothesis predicts no effect of lexical pre-activation. The Conceptual Level Hypothesis predicts that gestures are produced when the mapping from imagistic representations to propositional representations is complicated or when the information at the conceptual level requires additional computations. The Lexical Level Hypothesis predicts no effect of conceptual complexity.

To test these hypotheses, we analyzed a series of brief picture descriptions and compared when gestures were produced to when gestures were not produced. We then examined the results with respect to the predictions of these two hypotheses.

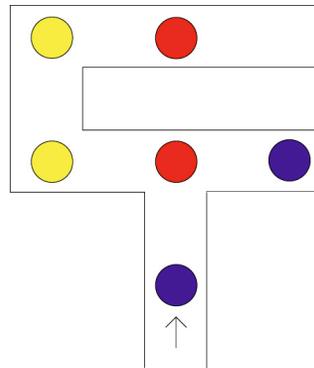
### **3. Experiment**

We presented 16 native speakers of Dutch with 16 abstract map-like images. The maps depicted segments of streets and intersections with destinations arranged along a path. Destinations were large colored circles positioned in the middle of the streets.<sup>1</sup> Images contained either five or six destinations, and half included a branching route. An example of the images used in this study is provided in (1).

---

<sup>1</sup> To facilitate memorization, destinations were limited to 3 colors: yellow, blue, and red.

(1) Sample image used to elicit path descriptions



Participants were seated across from their interlocutor with a visual block separating them. Participants were instructed to describe, from memory, a route that led past all the destinations in each image. Their description was supposed to be detailed enough to allow the non-visible listener to envision the picture and answer questions about it. Each picture was placed on a table in front of the participant for approximately 15 seconds. After this memorization period, the picture was removed and the participant began to describe the image. Participants were free to describe the routes in any way that was natural to them. They were not given any linguistic examples to bias their description strategy. The listener (who was in fact a conspirator) was instructed not to ask any specific questions that might bias the content of the descriptions. She was free, however, to ask the speaker to repeat portions or even the entire description of an image. Participants were video recorded from two angles to capture their gestural and linguistic behavior. The first and last four pictures for each participant were coded for speech and gestures.

### 3.1. Linguistic Data

For the purposes of conducting a quantitative analysis, several types of linguistic information were identified from the transcriptions of the participants' descriptions: directional information, destination information, landmark information, and overview information. To clarify these distinctions, consider the constructed sample description provided in (2) below. Content specifying a direction of movement, such the italicized portions of (2a) and (2d), was classified as directional information. References to the colored circles as the destination of a movement, such as the underlined portions of (2a) and (2d), were classified as destination information. References to circles as the goal of a movement were distinguished from references to circles as the source of a movement, such as the italicized portion of (2f). References to source information, as well as references to the characteristics of the road, as in the italicized portions of (2b) and (2e), were classified as landmarks, since they functioned as reference points within the image. Information about the general shape of the image or a sub-portion of the image,

*Does Gesture Help Processes of Speech Production?*

such as in (2d), and information about the relationship between multiple destinations, such as in (2g), were both classified as overview information.

- (2) a. *You walk straight ahead and you come across a yellow circle.*  
 b. *The yellow circle is in the center of an *intersection*,*  
 c. *just a plus-form-like intersection.*  
 d. *When you *turn left* you come across a red circle.*  
 e. *Now you go back to the center of the *intersection*, so the yellow circle.*  
 f. *From the *yellow circle* we walk straight ahead and you see a blue circle.*  
 g. *So you have three circles all in a row.*

To successfully complete the task, participants needed to minimally include direction and destination information. However, since they were given no examples of how to conduct their descriptions, they were free to include as much or as little information as they felt was needed. Thus, there was a wide amount of variation in the frequency with which non-essential landmark and overview information was included. Five participants generally included only the essential information. Five participants provided some landmarks in addition to the essential information. Six participants regularly included all four types of information in their descriptions. The average number of linguistic mentions for each of these types of linguistic information is provided in (3).

- (3) Average number (per picture) of linguistic references to directions, balls, and non-essential information

|              | Essential information |              | Non-essential information |          |
|--------------|-----------------------|--------------|---------------------------|----------|
|              | Directions            | Destinations | Landmarks                 | Overview |
| Participants | 5.1                   | 6.6          | 1.5                       | 0.3      |

Speakers were also free to adopt whatever spatial perspective they chose. They could describe the directions as though they were moving through the path (intrinsic perspective) or they could describe the objective directions as they were seen on the paper (deictic perspective or ‘bird’s eye view’, cf. Levelt 1987). Deictic speakers use directional terms like up, down, left, and right. In contrast, intrinsic speakers use straight ahead, left, and right. Since intrinsic speakers move through the image, what they see on the paper and what they say often conflict; if they have traveled in a circle, left may be right and right may be left. This mismatch between what is on the picture and what is said does not occur for deictic speakers. The constructed examples in (4a) and (4b), which are descriptions of a portion of the image in (1), demonstrate the differences between these two perspectives. We will return to this issue of perspective-taking later in the paper.

- (4) a. Deictic Perspective: You start at a blue circle, then you go up to a red circle. Next you go to the right and you see another blue circle. Now go back to the red circle where you just were and go to the left. You'll see a yellow circle. At the yellow circle, go up...
- b. Intrinsic Perspective: You start at a blue circle, then you go straight ahead until you come to a red circle. Next you go to the right and you see another blue circle. Now, go back to the red circle where you just were and then go straight ahead until you come to a yellow circle. At the yellow circle, you go to the right...

### 3.2. Gestural Behavior

Using the video recordings of the picture descriptions, we identified all gestures that were semantically co-expressive with the concurrent speech. For example, if the speaker said “you go to the right” and simultaneously pointed to the right, we classified that as a co-expressive gesture. Gestures were generally produced with one or both hands. They were generally small movements produced either in the participant’s lap or in front of the torso, close to the body. In addition to gestures produced with one or both hands, we also identified head movements that were co-expressive with the ongoing speech.

We found several different types of gestures, most corresponding to the different categories of linguistic information. The most common type of gestures were hand or head movements indicating the direction of movement in the description. Right and left were gestured most often, but up, down, and straight ahead also received many co-expressive gestures. The next most frequent type of gesture was the pointing gesture, which indicates the location of a destination point in the imaginary image created by the speaker in the gesture space. Sometimes, in addition to a pointing gesture in which the participant would draw a circle in the gesture space to indicate the colored circles that represented the destinations in the pictures. Other representational gestures (often produced as air drawings) depicted the overall shape of the image (e.g. an F-form), some component of the image (e.g. a T-intersection), or the relationship between two circles (e.g. one is directly above the other). The table in (5) shows the percentage of linguistic mentions that were produced with a co-expressive gesture for each type of information.

- (5) Average percent of linguistic mentions that were produced with or without a gesture for direction, destination, and non-essential information

|                   | Direction information | Destination information | Landmarks and overview |
|-------------------|-----------------------|-------------------------|------------------------|
| With a gesture    | 25 %                  | 8 %                     | 3 %                    |
| Without a gesture | 75 %                  | 92 %                    | 97 %                   |

## *Does Gesture Help Processes of Speech Production?*

On average, participants made 2.2 gestures per picture. However, there was also a large amount of variation between participants. Many participants produced virtually no gestures (N=5); others produced very few gestures, namely less than one gesture per picture (N=5); and others made many gestures, namely an average of five gestures per picture (N=6).

### **4. The Relationship Between Gesture and Speech Production**

In order to assess the stage at which speech and gesture interact, we conducted a series of analyses on the data obtained from this elicitation procedure aimed at uncovering any differences in gesture patterns. Each analysis targeted one of the predictions of the two major hypotheses under consideration: namely, the Lexical Level Hypothesis and the Conceptual Level Hypothesis. To test each hypothesis, we identified points within and between pictures as more or less challenging at either a lexical or conceptual level and then compared the number of gestures that were produced in the two cases.

Our interpretation of differences in gesture frequency is based on the assumption that when a particular process is taxed, or when processing at a given level is taxed, speakers try to reduce the processing load by producing any behavior that will ease the load on that process. Thus, if gestures function to aid lexical retrieval, differences in complexity at the conceptual level should not affect the gesture behavior. Rather, any differences in gestural behavior should be attributable to differences in lexical content or factors related to lexical retrieval. Taken the other way, differences in gesture behavior should be attributable to differences in lexical retrieval. For example, the second mention of a lexical item should receive fewer co-expressive gestures than the first mention of the same lexical item when they occur in adjacent utterances. In contrast, if gestures function to aid conceptual processing, then differences in conceptual planning or processing should affect gesture frequency, while differences in lexical content or lexical processing demands should have no effect. For example, portions of the pictures that were conceptually or computationally more complex should elicit more gestures than less challenging portions of the pictures.

#### **4.1. Did the Chance of Producing a Gesture Increase as the Computations at the Conceptual Level Increased?**

If gestures facilitate conceptual processes, then differences in processing or planning demands at the conceptual stage should correspond to differences in gesture frequency. To test this hypothesis, we identified conceptually challenging sub-portions of our pictures. Specifically, we chose the points in the pictures when the route branched in two directions. At these points, participants were required to take a number of additional steps at the conceptual level. First, the participant had to decide which path to travel first. Next, they had to remember the color of the 'choice point ball' so they could return to it. Finally, after describing the first branch, they had to return to the choice point and decide which way to go next. For intrinsic speakers, this requires recalculating the next direction since they are not

longer facing the same direction they would have been facing if they had chosen to travel the second branch first. Thus, describing the movements around a choice point is relatively taxing compared to describing the other deterministic movements within the picture.

We targeted the three movements around the choice point as conceptually most challenging: (i) the initial movement away from the choice point, (ii) the return to the choice point, and (iii) the final movement away from the choice point, after which the path is deterministic again. For each participant who produced at least one gesture in some picture, we calculated the percentage of directional gestures produced in this three-movement window compared to the percentage of directional gestures produced overall. These figures, averaged across participants, are presented in (6) below.

(6) Average gesture frequency at conceptually challenging sub-portions of descriptions compared to overall frequency

|                   | Directional terms<br>with a gesture | Directional terms<br>without a gesture |
|-------------------|-------------------------------------|----------------------------------------|
| At decision point | 48 %                                | 52 %                                   |
| Overall           | 33 %                                | 67 %                                   |

As can be seen, the percentage of directional gestures produced within the three-movement window around the choice point is considerably higher than the percentage of gestures produced overall. This suggests that gesture production increases as the conceptual difficulty of the task increases. Given that the lexical content for these moves is essentially the same as other, non-choice point moves, the differences in gesture frequency cannot be attributed to processes of lexical retrieval.

#### 4.2. Did Intrinsic Speakers Gesture More than Deictic Speakers?

Another test of the conceptual level hypothesis is to compare the percentage of gestures produced by intrinsic speakers to the percentage of gestures produced by deictic speakers. When deictic speakers provide the direction for a movement, they must examine the image in memory and identify the direction. In contrast, since intrinsic speakers move throughout the image, the direction of each movement is dependent on where they are in the image at that moment and what direction they are facing. Thus, they must compute the direction for each turn as they come to it. We argue that this is more complex a task, and this is supported by the fact that some participants demonstrate overt problems with the task. Often participants will pause when confronted with a particularly difficult calculation. Sometimes they will even turn themselves around in the chair to help visualize which direction they would be facing if they had actually moved through the picture. Thus, at the conceptual planning level, intrinsic perspective is more difficult than deictic

### *Does Gesture Help Processes of Speech Production?*

perspective. However, since the task remains the same, the words used to describe the movements are essentially the same.<sup>2</sup> The average numbers of directional gestures produced by intrinsic and deictic speakers, respectively, are shown in (7) below. As can be seen, intrinsic speakers produce many more directional gestures, on average, than deictic speakers, again supporting the Conceptual Level Hypothesis.

- (7) Average number of directional terms that were produced with and without a co-speech gesture by deictic and intrinsic speakers

|                           | Directional referents<br>with a gesture | Directional referents<br>without a gesture |
|---------------------------|-----------------------------------------|--------------------------------------------|
| Deictic speakers (N=5)    | 9.6                                     | 32.8                                       |
| Intrinsic speakers (N=11) | 12.2                                    | 25.8                                       |

Although it was not the case that adopting the intrinsic perspective automatically resulted in a large number of gestures, we did find that, on average, intrinsic speakers produced more directional gestures than deictic speakers. In contrast, their average number of linguistic mentions did not differ. Thus, while the linguistic content was essentially the same, their gesture behavior was different. As with the choice point data, these results cannot easily be interpreted as lexical retrieval facilitation.

We have shown two sources of evidence that complexity at the conceptual level increases the number of gestures produced. This seems to be strong evidence to support the Conceptual Level Hypothesis. However, the Conceptual Level Hypothesis and the Lexical Level Hypothesis need not be mutually exclusive. It is possible that gestures could facilitate processes at both levels. Therefore, our next analysis aimed to find some evidence that gestures facilitate lexical retrieval.

#### **4.3. Did the Chance of Producing a Gesture Decrease as Ease of Lexical Retrieval Increased?**

If gestures aid processes of lexical retrieval, then moments of relatively difficult retrieval should produce more gestures than moments of relatively easy retrieval. To investigate this prediction, we analyzed the gesture frequency on all lexical items that were repeated in close proximity with the same intended referent.<sup>3</sup> Models of word production and recognition have found that the activation of a word is facilitated if it is preceded by a related word. This facilitation is greatest when the preceding word is identical to the target word. Based on this well-established finding, we assume that the second mention of any word should be easier to retrieve than the first mention of the same word when they occur within

<sup>2</sup> Right and left are the same. Deictic speakers use up and down, while intrinsic speakers use straight instead.

<sup>3</sup> We would like to thank Dr. Gabriella Vigliocco for suggesting this analysis.

the same clause or in immediately adjacent clauses. In contrast, we assume that the conceptual processes that underlie the production of the first and second mentions are equivalent. While there may be some priming at the conceptual level for the single repeated concept, that priming occurs within the domain of the construction of an entirely new utterance and therefore should not have a discernible effect on the ease of conceptualization. Thus, in utterances like those in (8), the Lexical Level Hypothesis would predict more semantically co-expressive gestures to co-occur with the first mention of *road* than are found for the second mention, since retrieval of the second mention is already facilitated by the first.

- (8) a. ...and you take the road, the road goes again upwards.  
 b. ...you can go straight ahead or to the left. You go to the left.

An utterance was considered a repetition if the same lexical item was used to refer to the same referent either within the same clause or in the following clause. Repetitions occurred fairly frequently (N=75 pairs); however, only a few had co-expressive gestures on the first or second (or third) mention (N=24). The numbers of first and subsequent mentions of a given lexical item that occurred with a co-expressive gesture are presented in (9) below.

- (9) Number of semantically co-expressive gestures produced with the first mention or immediately following mention of the same lexical item with the same referent

|                                | Directions | Destinations | Non-essential | Total |
|--------------------------------|------------|--------------|---------------|-------|
| 1 <sup>st</sup> mention        | 4          | 2            | 3             | 9     |
| 2 <sup>nd</sup> primed mention | 3          | 5            | 7             | 15    |

As this table shows, there was no tendency to produce a gesture on a first mention rather than on a second or subsequent mention, suggesting that these gestures were not produced to facilitate lexical retrieval.<sup>4</sup> Gesture frequency did not decrease when a lexical item was primed due to repetition. This suggests that these gestures were not produced to facilitate lexical retrieval.

## 5. Discussion

Most models of speech production agree that conceptualization precedes linguistic formulation. The processes at the conceptual level are varied but include accessing visual, spatial, and encyclopedic information about the world and the ideas to be expressed. If a speaker wants to produce the word *rabbit* she activates many types

<sup>4</sup> For the purposes of this analysis we only investigated the six participants who produced more than one gesture per picture.

## *Does Gesture Help Processes of Speech Production?*

of information about the rabbit, including that the fact that it is an animal, has long ears, eats lettuce, etc. These conceptual and semantic features combine to activate a lemma representation corresponding to the word *rabbit*. This lemma representation is associated to morphosyntactic (and possibly also semantic) features of the word.

Given this architecture, there are two ways in which lexical retrieval can be facilitated by gestures. The first route, proposed by the Lexical Level Hypothesis, is a direct activation of the lexical representations by the gesture. According to this hypothesis, gestures prime lexical representations directly, bypassing conceptual level planning. The second possible route is indirect activation of the lexical representation mediated by the conceptual representations, as proposed by the Conceptual Level Hypothesis. According to this proposal, gestures function to activate conceptual level representations and/or aid in conceptual level processes. The representations at the conceptual level then feed activation down to the lexical representation. Given both of these possible accounts for how lexical retrieval might benefit from gestures, the evidence cited in support of the Lexical Level Hypothesis (e.g. that gesture prohibition adversely affects speech production) is naturally explained by both hypotheses.

Thus, to disentangle these two hypotheses of how gestures might facilitate speech production, we attempted to isolate components of our picture description task that were more or less challenging at one of the two processing levels. To investigate the issue of conceptual level facilitation, we contrasted gesture frequency for intrinsic and deictic speakers, arguing that the latter required more computations at the conceptual level than the former. As predicted, we found that the intrinsic speakers produced more gestures than the deictic speakers. Furthermore, we also found that when the demands of the description task were greater, specifically around the decision point, speakers also produced more gestures. Both of these findings support the claims of the Conceptual Level Hypothesis. In contrast, when we compared the number of gestures produced at the first or second mention of a repeated lexical item we found no difference in the gesture frequency. This is counter to the prediction of the Lexical Level Hypothesis.

As suggested earlier, the lexical and conceptual level hypotheses need not be mutually exclusive. It is possible that gesture has a direct facilitative effect at both the lexical and the conceptual levels. The results discussed in the literature are compatible with both hypotheses. Granting this, the results presented in this paper seem to support facilitation only at the conceptual level. However, it should be noted that the results presented here are numerical trends, not statistical significant differences. Furthermore, even finding a statistically significant difference using the correlational design presented above would not prove a causal link between increased processing load and gesture; rather, it would only be suggestive of a relationship at the conceptual level. In addition, it cannot be concluded with certainty that no facilitative effect is present at the lexical level, despite the fact that we found no evidence to support such a conclusion. Most of the lexical items used by the speakers in our study were relatively high frequency and common; they were also repeated over and over again across the sixteen pictures. Thus, it is possible

that this task is not optimally designed to reveal effects of lexical level facilitation. The challenge is to find a task that contrasts lexical properties of words while keeping conceptual properties constant. This is what we attempted to do in the present study. Our results clearly suggest that gestures have a facilitative function at the conceptual level, but whether they may also have a direct, unmediated facilitative function at the lexical level is still unclear despite the absence of evidence to date.

## References

- Alibali, Martha W., Sotaro Kita, and Amanda J. Young. 2000. Gesture and the Process of Speech Production: We Think, Therefore We Gesture. *Language and Cognitive Processes* 15:593-613.
- Butterworth, Brian. 1989. Lexical Access in Speech Production. In William Marslen-Wilson et al. (ed.), *Lexical Representation and Process*. Cambridge, MA: MIT Press.
- Butterworth, Brian, and Uri Hadar. 1989. Gesture, Speech and Computational Stages: A Reply to McNeill. *Psychological Review* 96:168-174.
- Cohen, Akiba. 1977. The Communicative Functions of Hand Illustration. *Journal of Communications* 27:54-63.
- de Ruiter, Jan-Peter. 1998. Gesture and Speech Production. Ph.D. diss., Max Planck Institute for Psycholinguistics, The Netherlands.
- Dell, Gary. 1986. A Spreading Activation Theory of Retrieval in Sentence Production. *Psychological Review* 93:283-321.
- Frick-Horbury, Donna, and Robert E. Guttentag. 1998. The Effects of Restricting Hand Gesture Production on Lexical Retrieval and Free Recall. *American Journal of Psychology* 111:43-62.
- Kempen, Gerard, and Pieter Huijbers. 1983. The Lexicalization Process in Sentence Production and Naming: Indirect Selection of Words. *Cognition* 14:185-209.
- Kendon, Adam. 1994. Do Gestures Communicate?: A Review. *Research on Language and Social Interaction* 27:175-200.
- Kita, Sotaro. 2000. How Representational Gestures Help Speaking. In D. McNeill (ed.), *Language and Gesture: Window into Thought and Action*. Cambridge, UK: Cambridge University Press.
- Krauss, Robert, Yihsiu Chen, and P. Chawla. 1996. Nonverbal Behavior and Nonverbal Communication: What Do Conversational Hand Gestures Tell Us? *Advances in Experimental Social Psychology* 28:389-450.
- Krauss, Robert, Yihsiu Chen, and R. Gottesman. 2000. Lexical Gestures and Lexical Access: A Process Model. In D. McNeill (ed.), *Language and Gesture: Window into Thought and Action*. Cambridge, UK: Cambridge University Press.

*Does Gesture Help Processes of Speech Production?*

- Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, Willem J. M., Ardi Roelofs, and Antje S. Meyer. 1999. A Theory of Lexical Access in Speech Production. *Behavioral and Brain Sciences* 22:1-75.
- Rauscher, Frances H., Robert Krauss, and Yihsui Chen. 1996. Gesture, Speech and Lexical Access: The Role of Lexical Movement in Speech Production. *Psychological Science* 7:226-231.
- Slobin, Dan. 1987. Thinking for Speaking. In J. Aske, N. Beery, L. Michaelis, and H. Filip (eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society, 435-444.

FR 4.7 Psycholinguistik  
Universität des Saarlandes  
Geb. 17.1, Room 1.16  
D-66041 Saarbrücken  
Germany

[melinger@coli.uni-sb.de](mailto:melinger@coli.uni-sb.de)

Max Planck Institute for Psycholinguistics  
Wundtlaan 1  
P.O. Box 310  
6500 AH Nijmegen  
The Netherlands

[kita@mpi.nl](mailto:kita@mpi.nl)



# What Do Speech-Gesture Mismatches Reveal about Speech and Gesture Integration? A Comparison of English and Turkish

ASLI ÖZYÜREK  
*Koç University*

## 0. Introduction

The semantic and temporal synchrony between speech and spontaneous hand gestures has been taken as evidence that speech and gestures are part of the same system (McNeill 1992, Kendon 1997). This paper questions the validity of this assumption through a cross-linguistic comparison.

Work by Talmy (1985), Slobin (1996), and others has shown that languages lexicalize the semantic components of spatial relations in different ways. The general question investigated in this paper is whether and how gestures synchronize *semantically* and *temporally* with the accompanying speech in languages where the semantic elements of a motion event are lexicalized differently—namely, in Turkish and English. This question is investigated in two studies.

The first study questions the semantic synchrony assumption. Do gestures represent semantic elements of a motion event in the same way in different languages, or does the representation in gestures vary from one language to another as the lexical and syntactic encoding of semantic elements vary? If the semantic synchrony assumption is true, then it is expected that gestures of speakers of different languages will vary with differences in lexicalization patterns of semantic elements in these languages.

The second study investigates the temporal synchrony assumption. Does the information in the gesture content temporally synchronize with the information in the accompanying speech content in different languages? If the temporal synchrony assumption is true, then what is expressed in gesture will temporally overlap with what is expressed in the co-temporal speech in different languages.

## 1. Study 1: Semantic Synchrony

With regard to expressing motion events, such as describing a ball rolling down a hill, English and Turkish differ from each other in the way they lexicalize manner and path elements of a motion event, as seen in (1).

(1) English and Turkish expressions of a motion event

**English:** Speakers can express manner and path components within one verbal clause.

|              |           |
|--------------|-----------|
| “rolls down” |           |
| V            | satellite |
| manner       | path      |

**Turkish:** Speakers have to use two different verbal clauses to express manner and path.

|                        |           |
|------------------------|-----------|
| “yuvarlan-arak iniyor” |           |
| V-roll-Conn            | V-descend |
| manner                 | path      |

In previous work (Özyürek and Kita 1999), we tested whether Turkish and English speakers’ gestures vary paralleling these differences in the lexicalization patterns.

### 1.1. Subjects

15 American English and 17 Turkish speakers participated in this study. All subjects were monolingual speakers.

### 1.2. Method

Each subject was asked to see and talk about an animated cartoon “Canary Row” (8 minutes). In the cartoon Sylvester the Cat attempts to catch Tweety Bird in different ways, each including a series of motion events.

### 1.3. Coding

One scene from the cartoon was selected for detailed analysis of speech and gesture. In this scene Sylvester swallows a bowling ball that Tweety Bird throws into his mouth and with the force of this bowling ball he rolls down the street and ends up in a bowling alley. The linguistic expressions and gestures were coded as follows.

Linguistic expressions: Verbal descriptions of this scene were coded for whether each speaker used (a) a verb + satellite construction, or (b) separate verbs to describe the manner and path components of the cat’s rolling down the hill.

Gestures: Speakers’ gestures that accompanied verbal expressions of this scene were categorized into three types.

*What Do Speech-Gesture Mismatches Reveal?*

- a) Manner-only gestures: Representing the manner of the motion event only (i.e. hand(s) or fingers rotate/wiggle without any trajectory component)
- b) Path-only gestures: Representing the path of the motion event only (i.e. hand(s) move along a lateral or sagittal trajectory without any rotation/wiggling of the hands or fingers)
- c) Manner-path conflated gestures: Representing both the path and the manner of motion simultaneously (i.e. hand(s) move along a lateral or sagittal trajectory while the hands or the fingers rotate/wiggle)

**1.4. Results**

**1.4.1. Speech**

English speakers used one verbal clause (e.g. *He rolls down*) to express both manner and path in the rolling event, whereas Turkish speakers used two verbal clauses (e.g. *He rolls and goes down the street*), as shown in Table 1. In the English sample, three speakers also used a manner-only verbal clause (e.g. *He is rolling*) and one speaker used a path-only clause (e.g. *He goes down*) in addition to their main verbal clause.

**Table 1.** Percentage number of subjects who expressed path and manner within one clause or in two clauses in the Turkish and English sample

| Number of subjects | One clause | Two separate clauses |
|--------------------|------------|----------------------|
| English (N=15)     | 100 %      | 0 %                  |
| Turkish (N=17)     | 0 %        | 100 %                |

**1.4.2. Gestures**

We also looked at whether Turkish and English speakers varied in terms of the way they used manner-only, path-only, and manner-path conflated gestures. First, more Turkish speakers than English speakers used manner-only gestures to describe the scene, as can be seen in Table 2.

**Table 2.** Percentage number of subjects who used manner-only gestures at least once in their repertoire of gestures

| Number of subjects | Used at least once | Never used |
|--------------------|--------------------|------------|
| English (N=15)     | 7 %                | 93 %       |
| Turkish (N=17)     | 50 %               | 50 %       |

Second, more Turkish speakers than English speakers used path-only gestures to describe the scene, as seen in Table 3.

**Table 3.** Percentage number of subjects who used path-only gestures at least once in their repertoire of gestures

| Number of subjects | Used at least once | Never used |
|--------------------|--------------------|------------|
| English (N=15)     | 43 %               | 57 %       |
| Turkish (N=17)     | 69 %               | 31 %       |

Lastly, there was no difference between the number of Turkish and English speakers who used manner-path conflated gestures, as shown in Table 4.

**Table 4.** Percentage number of subjects who used manner-path conflated gestures at least once in their repertoire of gestures

| Number of subjects | Used at least once | Never used |
|--------------------|--------------------|------------|
| English (N=15)     | 71%                | 29%        |
| Turkish (N=17)     | 69%                | 31%        |

### 1.5. Summary and Conclusion

The way Turkish and English speakers used their gestures to represent the elements of a motion event paralleled the differences in their lexicalization of semantic elements. Turkish speakers used more manner-only and path-only gestures, which paralleled the fact that they used separate verbal clauses to describe both manner and path. In contrast, English speakers mostly used manner-path conflated gestures, which paralleled the fact that they can express both elements within one verbal clause; few of them used manner-only or path-only gestures.

These findings show that gestures also differ when the linguistic encoding possibilities vary in different languages. This provides evidence for the view that gestures and speech have semantic synchrony across different languages.

## 2. Study 2: Temporal Synchrony

In this section I investigate whether speakers of Turkish and English also temporally coordinate the content of their gestures (i.e. manner-only gestures) with the content of their speech (i.e. manner-only clauses). That is, does the content of gestures match the speech content they synchronise with temporally? There could be two possibilities:

- a) Match: Gesture content matches the co-temporal speech content. In this case there is semantic overlap (partial or total) between what is expressed in gesture and the co-temporal speech (e.g. Speech: *He goes down the street*, Gesture: path-only gesture).



- c. Speech: *yuvarlana yuvarlana*  
rolling rolling  
Gesture: path-only MISMATCH
- d. Speech: *sokaktan*  
on the street  
Gesture: path-only MATCH
- e. Speech: *gidiyo*  
goes  
Gesture: path-only MATCH

Translation: ‘Ball somehow, hopping hopping, rolling rolling goes along the street’

These results show that the temporal synchrony assumption does not hold all the time for Turkish speakers, as we see in (a) and (c) above.

However, even though Turkish speakers do not frequently synchronize their gesture content with the exact temporal speech content (i.e. within one gesture-speech combination unit), they might be trying to synchronise the information content in their speech and gesture at the *sentence level*. In order to test this possibility, the information content in the whole sentence used to describe the motion event scene was compared with the information content of the co-temporal gestures. For example, in the Turkish case above the information content in the whole sentence was compared to the information content revealed in the five gestures that overlapped with the whole sentence. In this example, if we take the whole sentence into consideration, the information content in speech and gesture match. The analysis in Table 6 shows that the content of gestures and the co-temporal accompanying speech content match for both speakers of Turkish and English at the sentence level.

**Table 6.** Percentages of match and mismatch between speech and gesture at the sentence level

| Number of sentences | Match between speech and gesture | Mismatch between speech and gesture |
|---------------------|----------------------------------|-------------------------------------|
| English (N=22)      | 100 %                            | 0 %                                 |
| Turkish (N=17)      | 100 %                            | 0 %                                 |

## 2.2. Summary and Conclusion

Cross-linguistic comparison showed that in languages where lexical encoding of semantic elements is different, the temporal synchrony between speech and gesture does not always hold. That is, what is expressed in gesture and the exact co-temporal speech content do not always match in different languages. However, it is possible that in different languages the temporal synchrony between speech

### *What Do Speech-Gesture Mismatches Reveal?*

and gesture holds at different levels (e.g. the sentence). For example, what is at stake for a Turkish speaker is whether the information in the gesture content overlaps with the information expressed in the speech in the whole sentence, whereas for an English speaker the match is between what is in the gesture and the exact co-temporal speech content.

### **3. General Conclusion and Implications**

In this study, the assumption that speech and gesture have semantic and temporal synchrony was tested by comparing speech and gestures in two languages where the mapping between lexical and semantic elements is different—namely, in Turkish and English.

Study 1 showed that the semantic synchrony assumption holds across languages. Speakers of different languages use different gestures with different lexicalization patterns of semantic elements even though they describe the same motion event. Study 2, however, showed that the temporal synchrony assumption does not always hold for speakers of different languages. That is, what is expressed in gesture and the content of the exact co-temporal speech content do not always match.

The mismatches found in the Turkish sample show that the temporal synchrony assumption (McNeill 1992) between speech and gesture should be modified in ways that can cover cross-linguistic differences. It is possible that the temporal synchronization of the content in speech and the content in gesture is regulated differently during speech and gesture production in different languages. These findings also have implications for the claim that speakers of different languages plan their thinking for speaking in different ways as revealed by different temporal synchrony relationships between their speech and gestures (Slobin 1996, Kita 2000).

### **References**

- Kita, S. 2000. How Representational Gestures Help Speaking. In D. McNeill (ed.), *Language and Gesture*. Cambridge: Cambridge University Press, 162-185.
- McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press
- Özyürek, A., and S. Kita. 1999. Expressing Manner and Path in English and Turkish: Differences in Speech, Gesture and Conceptualization. In M. Hahn and S. Stoness (eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*. London: Lawrence Erlbaum Publishers, 507-512.
- Slobin, D. 1996. Two Ways to Travel: Verbs of Motion in English and Spanish. In M. Shibatani and S. A. Thompson (eds.), *Essays in Semantics*. Oxford: Oxford University Press, 195-217.

Asli Özyürek

Talmy, L. 1985. Lexicalization Patterns: Semantic Structure in Lexical Forms. In T. Shopen (ed.), *Language Typology and Syntactic Description, Vol. III: Grammatical Categories and the Lexicon*. Cambridge: Cambridge University Press, 57-149.

Department of Psychology  
Koç University  
Rumeli Feneri yolu, 819100  
Sarıyer, Istanbul  
Turkey

aozyurek@ku.edu.tr

## Gestures and Self-Monitoring in Speech Production

MANDANA SEYFEDDINIPUR and SOTARO KITA  
*Max Planck Institute for Psycholinguistics*

### 0. Introduction

Gesture and speech are semantically and temporally tightly coordinated. Gestures are pre-positioned temporally to the lexical affiliate with which they share semantic and/or pragmatic content (Kendon 1972, McNeill 1992, Schegloff 1984, Morrel-Samuels and Krauss 1992). The specific timing and semantic relation of speech and gesture has led to the view that gesture can serve as a “window” into mental processes underlying speech production (McNeill 1992).

In this paper we aim to gain insight into how speakers monitor their own speech by observing accompanying gestures. In natural speech, disfluencies are ubiquitous and they come in various forms. The flow of speech is interrupted by pauses and filled pauses (*uhm*), speakers search for specific words, they have trouble in the articulation of a word, and speakers also correct themselves. Self-corrections consist of different processes. Speech has to be monitored for appropriateness and for correctness. If an error is detected, the speech stream has to be interrupted, the repair has to be planned, and it then has to be finally executed.

The tight coordination between speech and gesture has led to the conclusion that (at least) at the conceptual level, speech and gesture production are closely interrelated. Furthermore, it has been argued that self-monitoring of speech is a conceptual level process (as opposed to a formulational level process, cf. Levelt 1989). Thus, it can be expected that gesture is sensitive to speech disfluency. By utilizing the specific timing relation of speech and gesture, we test two views regarding how speakers monitor their own speech.

We will investigate this issue on the basis of a corpus of disfluencies comprising descriptions of houses and apartments. The description of living spaces has proven to be a useful task for eliciting various kinds of gestures and speech disfluencies. The speaker has to transform three-dimensional space into the linear structure of speech. In addition, the speaker has to choose the appropriate words and constructions in order to convey the selected and linearized spatial information in a comprehensible way (Ullmer-Ehrich 1982). These difficulties result in a high number of disfluencies of different kinds and in a considerable use of gesture.

### 1. Monitoring Theories

Speakers monitor their own delivery constantly. They control their delivery such that what is going to be said is what they had intended. More specifically, they control for the appropriateness of selected words, and they check for errors (for details of foci of monitoring, see Levelt 1983, 1989). If an inappropriateness is detected, the speaker interrupts his speech stream and repairs the erroneous or inappropriate utterance. This whole process consists of four components: monitoring of speech, error detection, self-interruption, and self-correction. Various psycholinguistic theoretical accounts of speech monitoring and error detection have been proposed (for a review, see Postma 2000). Two of these accounts will be tested in this paper.

The INTERRUPTION-UPON-DETECTION HYPOTHESIS states that the speech stream is interrupted as soon as an error is detected. This is expressed in the Main Interruption Rule: “Stop the flow of speech immediately upon detecting trouble” (Levelt 1983, 1989; Nootboom 1980). After the interruption of speech, the planning for reformulation takes place.

The rationale behind the Main Interruption Rule is that linguistic structures are ignored in interruption. Levelt’s (1983) analysis showed that speakers interrupted their speech stream at any point in the delivery. They did not attend to any linguistic boundaries like syllables, words, or phrase boundaries. One exception is that speakers tended to complete non-erroneous words, i.e. neutral or merely inappropriate ones. This led to the refinement of the model such that the Main Interruption Rule only applies to cases of immediate detection of erroneous words.

The DELAYED-INTERRUPTION-FOR-PLANNING HYPOTHESIS suggests that even if an error is detected, the speaker does not interrupt his flow of speech immediately (Blackmer and Mitton 1991, Fox Tree and Clark 1997, Clark and Wasow 1998). Upon detection of an error the speaker will start the replanning and interrupt when the repair is ready to a certain degree or the speaker has run out of what can be uttered without further conceptual processing.

Blackmer and Mitton (1991) based their hypothesis on the analysis of the temporal characteristics of self-repairs in spontaneous speech. They observed that time intervals between the interruption point and resumption of speech were sometimes shorter than predicted by Levelt’s Main Interruption Rule. According to the Main Interruption Rule, the replanning takes place only after the interruption. This implies that there has to be a time interval of some length before the resumption can take place. However, Blackmer and Mitton found instances where the suspension point was immediately followed by the correction, without any pause in between. Their results imply that the planning of the correction can take place while speaking is in progress and not only after suspension. Fox Tree and Clark (1997) came to a similar conclusion, but with a rather different type of evidence. They conducted a corpus study on the occurrence of the two pronunciation variants of the English article *the* (*thuh* with the reduced vowel schwa, and *thiy* with a non-reduced vowel). They found that 81% of the instances of *thiy* were followed by a suspension of speech. This suggests that speakers detected the

problem at some interval before suspending speech. By knowing in advance that they were going to suspend, the location of suspension (after *the*) and the type of suspension (the pronunciation of the variant *thiy*) is planned.

Taking the temporal and semantic interlocking of gesture and speech into account, the two theoretical approaches make different predictions concerning the gestural behavior. The Interruption-Upon-Detection Hypothesis predicts that any effect on gesture should be simultaneous with or follow the speech suspension. There should not be any effect on gesture before the actual speech suspension. This prediction is based on two assumptions: (i) when an error is detected, a stop signal is sent to both production modalities simultaneously (for an account of the suspension of speech and gesture production, see de Ruiter 2000), and (ii) it takes longer to suspend a gesture than speech because heavier mass has to be stopped in gesture.

In the case of the Delayed-Interruption-for-Planning Hypothesis, an effect on gesture can occur even before the moment of speech suspension due to the lag between error detection and speech suspension. When speakers have detected an error or have anticipated trouble, they start to plan how to resume right away and at the same time suspend the gestural movement. In the meantime, they go on speaking until the repair is ready up to a certain point or they have run out of formulated words. Consequently, gesture can stop before speech stops.

These predictions are tested by investigating the temporal relationship between different phases of self-repair and movement phases of gesture, which will be defined in the following sections.

### 3. Structural Components of Gesture and Speech Disfluencies

#### 3.1. Disfluency Structure

A speech disfluency can be divided into different phases following Clark's (1996) disruption schema:

| Suspension<br>Point | Resumption<br>Point         |
|---------------------|-----------------------------|
| “On the right       | uhm   on the left side....” |
| Original Delivery   | Hiatus Resumed Delivery     |

The first phase is the original delivery. The speaker monitors his internal speech for appropriateness and correctness (for a detailed description of foci of monitoring, see Levelt 1983). If an error is detected, the original delivery is disrupted. In the above example the original delivery is suspended at the word *right*. After the interruption a time interval (the hiatus) follows where speakers pause or utter filled pauses (e.g. *uhm*, *uh*) or so-called editing terms like *well*, *I think*, and *I mean*. The hiatus is seen as the phase where internal reformulation processes take place. The hiatus ends at the resumption point where the speaker resumes his delivery.

### 3.2. Gesture Structure

Gestures can be segmented into qualitatively different movement phases (Kendon 1972, 1980; McNeill 1992; Kita et al. 1998). The segmentation and identification of movement phases can be based purely on dynamic aspects of the hand/arm movement (as in Kita et al. 1998). In the preparatory movement phase the hands move from a resting position in order to prepare for the forcefully executed part, the stroke. The preparation phase can also be followed by a static phase, where the hands are held still in the initial position. This pre-stroke hold is then released by the stroke. The stroke phase is the semiotic and dynamic nucleus of the gesture. The stroke typically displays the meaning of the gesture. In the stroke the most force is exerted as compared to the neighboring phases. Also after this phase a static phase might follow, which is called the post-stroke hold. A gestural unit ends when the hands retract back into resting position, e.g. on the lap.

Preparation  $\Rightarrow$  Hold  $\Rightarrow$  **Stroke**  $\Rightarrow$  Hold  $\Rightarrow$  Retraction

Of the described gestural phases, only the stroke is obligatory. Note that in natural conversation one can observe a succession of strokes without the hands going into a hold or being retracted after each stroke.

## 4. Method

### 4.1. Stop Shifts and Start Shifts

In the analysis of the gestural movement pattern, we focused on the transition from one phase to another. Analogous to speech suspension and speech resumption, we distinguish two different types of phase shifts: a stop shift and a start shift. In a stop shift, an ongoing gestural unit/movement phase is suspended. In a start shift, a new dynamic gestural movement phase is initiated. These are described in more detail below.

**Stop shift:** an ongoing gestural movement is suspended or not completed.

- Shift of a dynamic phase into a static phase: an ongoing gestural movement phase (preparation/stroke) is suspended by going into a hold or by being retracted back into resting position.
- Shift of a dynamic phase into a new dynamic phase: a gesture gets suspended before being completed, e.g. a preparation phase is not followed by a stroke, for which the hands were preparing, but is followed by another preparation for the same or a different gesture.
- A dynamic phase is interrupted: a preparation or a stroke phase is prematurely truncated before the phase itself is terminated by a sudden abrupt halt or a sudden change in movement direction. In this case we classified the phase shift as a stop shift no matter what followed.

**Start shift:** a new gestural movement is started.

- Shift from a static phase into a dynamic phase: hands that are held still start a new preparation/stroke phase.
- A preparation phase is not followed by a stroke, but by a new preparation phase.
- An interrupted movement phase is followed by a new movement phase (preparation/stroke).

#### **4.2. Data**

The corpus consisted of six videotaped semi-natural conversations. Six native German speakers (four women, two men) were asked to describe houses and apartments they grew up in or had lived in for a longer period to a listener. Each session lasted 30-40 minutes. Nine minutes of the description from each speaker was transcribed. The speech data was coded for suspension points, hiatus length, and resumption points. The gestural movement phases were coded in terms of phase transitions. The temporal values were determined by a frame-by-frame microanalysis (1 frame = 40 ms). The six speakers produced a total of 582 disfluencies, of which 267 were overt repairs.<sup>1</sup> 191 overt repairs were accompanied by gestures, and 76 were not.

#### **4.3. Analysis**

We selected all utterances containing a repair that was accompanied by gestures (N=191). One speaker was excluded from the analysis because she did not provide sufficient data points. We analyzed the occurrences of stop shifts around suspension points and the occurrences of start shifts around resumption points. In order to ensure that the observations were independent from each other, we selected all repairs (i.e. the whole disfluency unit including suspension, hiatus, and resumption) that were at least two seconds apart from each other. We chose a time window of one second to each side of the suspension/resumption point and counted the number of start and stop shifts for every 160 ms slot within the window.

### **5. Results**

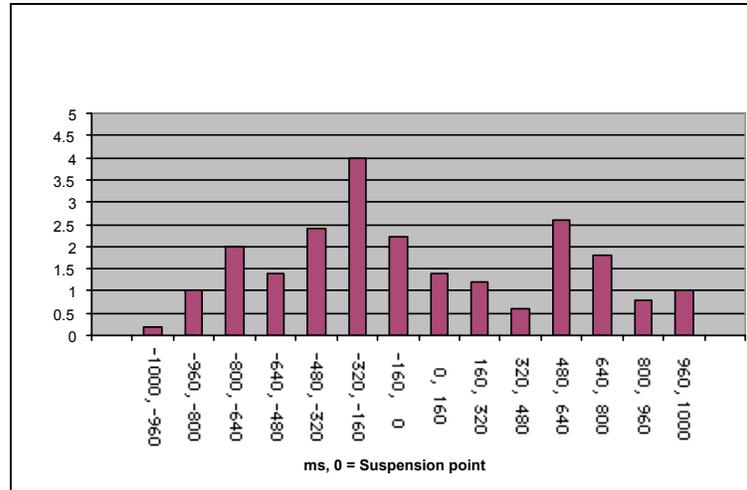
#### **5.1. Stop Shifts around the Suspension Point**

Figure 1 presents the frequency of stop shifts around the speech suspension point (averaged over five speakers). The one-second window before and after the speech suspension point is divided into 160 ms intervals (0 ms = suspension point). Each bar shows the average frequency of stop shifts for a given time interval.

---

<sup>1</sup> Following Levelt (1983), we distinguish between overt and covert repairs. In a covert repair, the resumption is a continuation of the original delivery without any alternation (e.g. *the living room was, uhm, on the left side*). In contrast, an overt repair is an instance of a disfluency where in the resumption an element is altered (e.g. *the living room was, uhm, the dining room was on the left*).

**Figure 1.** Average frequency of stop shifts around suspension points

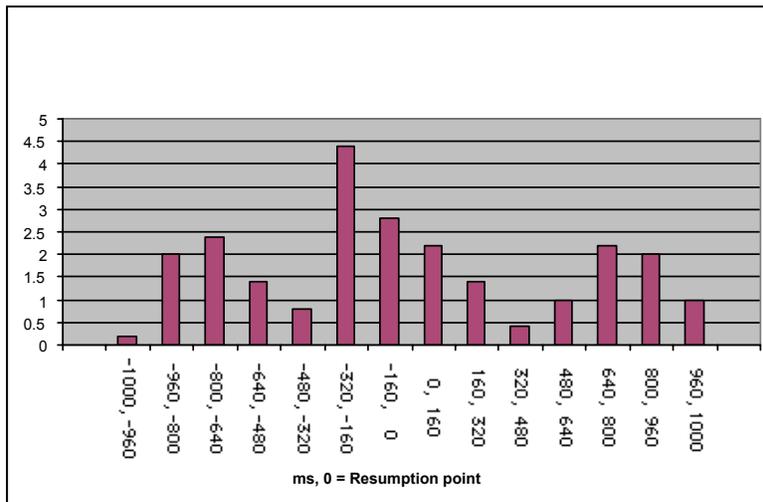


As is evident from Figure 1, gesture stops before speech stops. The most common time interval in which stop shifts occur is from -320 to -160 ms.

**5.2. Start Shifts around the Resumption Point**

Figure 2 presents the frequency of start shifts around the speech resumption point (averaged over five speakers). The one-second window before and after the speech resumption point is divided into 160 ms intervals (0 ms = resumption point). Each bar shows the average frequency of start shifts for a given time interval.

**Figure 2.** Average frequency of start shifts around resumption points



As is evident from Figure 2, gesture starts before speech starts. The most common time interval in which start shifts occur is from -320 to -160 ms.

## **6. Discussion**

The above results show that gesture is highly sensitive to speech disfluencies. When speech is suspended and then resumed, gesture is also suspended and then resumed. Suspension and resumption in the two modalities are temporally coordinated in a systematic way. This suggests a highly interactive planning process that is involved in the production of both modalities.

Gesture is suspended prior to speech suspension. This suggests that gesture can be seen as an indicator of an upcoming interruption in speech. The gestural foreshadowing of speech suspension suggests that speakers are already aware that there is or will be trouble, but they do not interrupt speech right away. This is predicted by the Delayed-Interruption-for-Planning Hypothesis, according to which speakers continue speaking after error detection. They start planning for the resumption already before the speech suspension and disrupt their delivery when the repair is ready to a certain degree or they have run out of words that can be formulated without further conceptual planning. The above result also indicates that at least some utterances are interrupted in the way not predicted by the Interruption-Upon-Detection Hypothesis, according to which gesture should be interrupted simultaneously with or even after speech suspension.

However, these two hypotheses are not mutually exclusive. A speaker may interrupt his/her speech in different ways depending on various contextual factors. For example, in order to avoid losing the floor, one might delay suspension of speech. At the same time, in order not to mislead the interlocutor, one might suspend and repair the error as soon as possible. The speaker has to always evaluate advantages and disadvantages of speech suspension at a given moment. The timing of a speaker's interruption of his/her speech may be determined by a moment-by-moment balance among competing factors like comprehensibility and floor-keeping.

There is an emerging view in the literature that speech interruption is not a reflex-like reaction to error detection, but a choice the speaker makes based on, for example, the abovementioned factors (Blackmer and Mitton 1991, Fox Tree and Clark 1997, Clark and Wasow 1998). This study provides novel converging evidence for this idea by using speech-accompanying gesture as a window into the speaker's mind.

## **References**

- Blackmer, Elizabeth R., and Janet L. Mitton. 1991. Theories of Monitoring and the Timing of Repairs in Spontaneous Speech. *Cognition* 39:173-194.
- Clark, Herbert. 1996. *Using Language*. Cambridge: Cambridge University Press.

- Clark, Herbert, and Thomas Wasow. 1998. Repeating Words in Spontaneous Speech. *Cognitive Psychology* 37:201-242.
- de Ruiter, Jan-Peter. 2000. The Production of Gesture and Speech. In D. McNeill (ed.), *Language and Gesture*. Cambridge: Cambridge University Press, 284-311.
- Fox Tree, Jean E., and Herbert Clark. 1997. Pronouncing “the” as “thee” to Signal Problems in Speaking. *Cognition* 62:151-167.
- Kendon, Adam. 1972. Some Relationships between Body Motion and Speech. In A. Siegman and B. Pope (eds.), *Studies in Dyadic Communication*. New York: Pergamon Press, 177-210.
- Kendon, Adam. 1980. Gesticulation and Speech: Two Aspects of the Process of Utterance. In M. R. Key (ed.), *The Relationship of Verbal and Nonverbal Communication*. The Hague: Mouton, 207-227.
- Kita, Sotaro, Ingeborg van Gijn, and Harry van der Hulst. 1998. Movement Phases in Signs and Co-Speech Gestures, and Their Transcription by Human Coders. In I. Wachsmuth and M. Froehlich (eds.), *Proceedings of the International Gesture Workshop: Gesture and Sign Language in Human-Computer Interaction*. Bielefeld, Germany, 17-19 September 1997. Berlin: Springer, 23-35.
- Levelt, Willem J. M. 1983. Monitoring and Self-Repair in Speech. *Cognition* 14:41-104.
- Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- McNeill, David. 1992. *Hand and Mind*. Chicago: University of Chicago Press.
- Morrel-Samuels, Palmer, and Robert M. Krauss. 1992. Word-Familiarity Predicts Temporal Asynchrony of Hand Gestures and Speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:615-622.
- Nooteboom, Sieb. 1980. Speaking and Unspeaking. Detection and Correction of Phonological and Lexical Errors in Spontaneous Speech. In V. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, 87-95.
- Postma, Albert. 2000. Detections of Errors during Speech Production: A Review of Speech Monitoring Models. *Cognition* 77:97-131.
- Shriberg, Liz. 1994. Preliminaries to a Theory of Speech Disfluencies. Ph.D. diss., University of California, Berkeley.

Max Planck Institute for Psycholinguistics  
Wundtlaan 1  
P.O. Box 310  
6500 AH Nijmegen  
The Netherlands

mandsey@mpi.nl  
kita@mpi.nl

# Comparing the Informational Content of Speech/Gesture and Sign Language

SARAH TAUB and PILAR PIÑAR  
*Gallaudet University*

## 0. Introduction

Linguists have always wanted to compare how languages express information. Do users of different languages communicate essentially the same information, or does language play a critical role in shaping what is included and excluded from a narrative? The “null hypothesis” or simplest assumption might be that humans in general attend to and express the same conceptual elements universally. This paper presents a research methodology that should allow us to evaluate that hypothesis. Earlier work has suggested that the hypothesis fails—that spoken languages differ to a large degree in what is expressed, and that sign languages draw on different resources and may express more information (in certain domains) than spoken languages. Taking spatial motion events as a test case, this paper argues that it is not appropriate to compare languages based on speech alone, or to compare speech to sign language; rather, the speech/gesture complex produced in discourse and narrative is analogous to sign language and constitutes the appropriate level for cross-linguistic comparison. If speech/gesture is taken as the level of comparison, we predict more universal trends in expression of information. The technique presented in this paper should allow collection of quantitative data over a large range of languages.

Talmy (1985) provided an analysis of spatial events upon which much subsequent typological research has been based. He divided motion events into such categories as Figure, Ground, Motion, Path, and Manner, and then classified languages as to how they typically expressed each class of information. In particular, *verb-framed* languages such as Spanish express Path information through verbs, and *satellite-framed* languages such as English express Path through other members of the verb complex (“satellites”). As examples, consider (1), a typical English sentence in which a bird’s Path through a window is described with a verb particle and prepositional phrase, and (2), a typical Spanish sentence where the same Path is described with a verb.

- (1) It flew out of the window.
- (2) Salió por la ventana.  
went-out:3SG through the window

Talmy noted that verb-framed languages are less likely to provide complex specifications of Path, as they would require a long series of verbs and thus many clauses (making the discourse feel “over-detailed”); as Manner requires an extra verb or gerund in these languages, it is often left out as well. On the other hand, satellite-framed manner-type languages such as English incorporate Manner easily in the main verb, then typically present long series of Path elements via several satellites. (3) demonstrates the use of a second verb in Spanish to specify Manner, and (4) demonstrates the use of multiple verb particles and prepositions in English to specify complex Paths.

- (3) Salió volando por la ventana.  
went-out:3SG flying through the window
- (4) It flew right back out the window, across the street, and into the forest.

Slobin’s (1996) studies of Spanish and English written narratives indicate that differences between the two languages (exemplars of the verb-framed and satellite-framed types) persist over a longer time period than the individual clause. That is, Spanish narratives consistently present fewer explicit Manner and Path elements than English narratives; they do not “catch up” in the course of the story. This analysis, however, considers only written texts and not oral narrative or discourse. Studies show (e.g. Kendon 1986, McNeill 1992) that the gesture that accompanies speech also contains conceptual elements (though it is a matter of hot debate whether such gesture is intended to communicate those elements).

Galvan and Taub (forthcoming) used a similar method of counting conceptual elements to compare the information expressed in American Sign Language and English narratives. Narratives were elicited by the same stimulus (the wordless storybook *Frog, Where Are You?*) and so could be compared more easily. They found that in nearly every category (Ground elements being the only exception), ASL storytellers included more conceptual information; in general, the signed narratives relied on ASL’s highly iconic grammatical resources to compress a huge amount of spatial information into each sentence. But again in this study, analysis of the English narratives looked only at the speech produced by the subject; there was no record of gesture or prosody.

Many people have noted similarities between sign languages and speakers’ gestures; many others have noted differences (the debate has a political edge, since when ASL was thought to be “only gesture”, deaf signers were thought to be without language and mentally impaired). Liddell (2000) has created a precise explanation for this phenomenon based on conceptual blending. According to

## *Comparing Speech/Gesture and Sign Language*

Liddell, both signers and speaking gesturers create a “blend” (Fauconnier and Turner 1996) between an imagined mental space and Real space (i.e. their conceptual model of the space in which they stand). In the blend, imagined entities are conceived of as located in the space around the person. Iconic and deictic gestures may be aimed toward those entities, or may track their progress through space.

For spoken languages, these gestures are loosely integrated with the speech signal: the emphatic stroke phase of the gesture occurs precisely with a spoken word or words that relates to its meaning (McNeill 1992), but the preparatory and refractory phases may vary in timing (indeed, McNeill notes that the precise timing of these other phases may reveal details of the speaker’s thought processes). For sign languages, on the other hand, these gestures are tightly constrained to unify with lexical and grammatical elements. Indeed, certain signs (e.g. pronouns, some verbs, classifier signs) are ungrammatical if they do not contain a gestural element. For one example, ASL pronouns must be placed in signing space so as to indicate the entity to which they refer; for another, verbs such as GIVE must indicate their subject or Source argument and object or Goal argument by moving from a space associated with the Source to a space associated with the Goal. We might describe these gestural elements of sign languages as “grammaticalized gesture”.

If Liddell’s proposal is correct, then sign language is more fully analogous to the speech/gesture complex than to speech alone. If co-speech gesture presents additional conceptual information, then comparison of speech/gesture across spoken languages may also prove more fruitful than comparison of speech alone. Elaborating on the null hypothesis above, if gesture is taken into consideration, cross-linguistically and cross-modally, in comparable narratives we may expect to see the following:

- A. Similar amounts of conceptual information
- B.
  - i. Similar types of information in speech and lexical sign elements
  - ii. Similar types of information in gesture and gestural sign elements
- C. Similar rates of information presentation

Given the research of Talmy and Slobin, we may hypothesize more specifically the following:

- D. Spanish uses gesture to add more Path and Manner information.

### **1. Methods**

The data presented in this paper are the first report on a larger ongoing typological study of English, Spanish, and ASL. This pilot study looked at two native users of each language. Each subject was paired with another native user of

his/her language. Subjects watched animated cartoons involving the adventures of a cat and a bird; they were instructed to tell what happened in the cartoon to their partner clearly enough so that the partner could then tell the story to a third person. All narratives were videotaped. Partners' narratives were recorded but not analyzed in this pilot study.

In the particular scene analyzed here, the cat and the bird are in high-rise apartments across the street from each other. The cat has been studying the bird through their windows. The cat swings from his window to the bird's window on a rope, but misses the window, crashes into the wall, and falls to the street below.

The researchers developed a list of conceptual elements present in the cartoon, including potential Figure, Ground, Path, Manner, and Instrument elements. The analysis compiled: (i) length of time to tell the scene, (ii) total number of conceptual elements expressed, (iii) number of elements expressed through speech or lexical sign elements, and (iv) number of elements expressed through gesture or gestural sign elements. Lexical and gestural sign elements were distinguished on the following criterion: if the element (e.g. hand shape, movement, location, complete sign) was conventionalized in the lexicon of ASL, then it was counted as lexical; if it was not, then it was counted as gestural. A number of signs contained both lexical and gestural elements.

## 2. Results

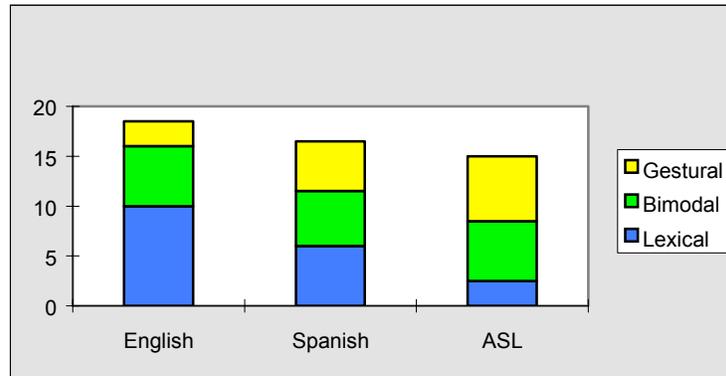
Table 1 summarizes the raw data for total conceptual items expressed by each subject. "Lexical" items are expressed through speech or lexical sign elements only; "gestural" items are expressed through gesture or gestural sign elements only; and "bimodal" items are expressed through both modes. Figure 1 gives the same information in bar graph form, with the two subjects' data averaged for each language.

**Table 1.** Data for each subject

| Language  | Total items | Lexical items | Bimodal items | Gestural items | Time (sec) | Items/second |
|-----------|-------------|---------------|---------------|----------------|------------|--------------|
| English 1 | 19          | 9             | 7             | 3              | 31         | 0.61         |
| English 2 | 18          | 11            | 5             | 2              | 24         | 0.75         |
| Spanish 1 | 15          | 5             | 6             | 4              | 6          | 2.5          |
| Spanish 2 | 18          | 7             | 5             | 6              | 6          | 3.0          |
| ASL 1     | 17          | 4             | 6             | 7              | 10         | 1.7          |
| ASL 2     | 13          | 1             | 6             | 6              | 9          | 1.4          |

*Comparing Speech/Gesture and Sign Language*

**Figure 1.** Total conceptual items



Approximately the same amount of information was expressed by each narrator, though the amount of time taken to tell the story varied widely. English subjects expressed more information through speech than through gesture, Spanish subjects expressed approximately the same amount through each modality, and ASL subjects expressed more information through gestural elements than through lexical elements. Without gesture, Spanish subjects produced only 11 and 12 items per story, and English subjects only 16 per story; with gesture, the totals rise to 15 and 18 for Spanish and 19 and 18 for English.

One Spanish speaker used a heavily “gesture-focused” style: the narrative is not comprehensible through the speech alone but requires gestural information. He referred explicitly to his gestures (e.g. “like this” or “here”) and used vocal gestures as well (e.g. “ñññ” accompanied the manual gesture representing the cat holding the rope and swinging across, “pom” accompanied the manual gesture representing the cat hitting the wall).

Tables 2-4 summarize the results for the major conceptual categories: Figure/Ground (lumped together here, since both represent “nominal” items), Path, and Manner. In these tables, the two subjects’ results are averaged for each language. Because the number of data points is so small, no statistical analyses were performed, but we observe that (just as in Table 1) the spread between the two subjects’ numbers is quite small. This information is also presented in bar graph form in Figures 2-4 below.

**Table 2.** Figure/Ground conceptual items

| Language | Lexical Items | Bimodal Items | Gestural Items |
|----------|---------------|---------------|----------------|
| English  | 4.5           | 0.5           | 0              |
| Spanish  | 1.5           | 1.5           | 1              |
| ASL      | 2             | 2             | 0              |

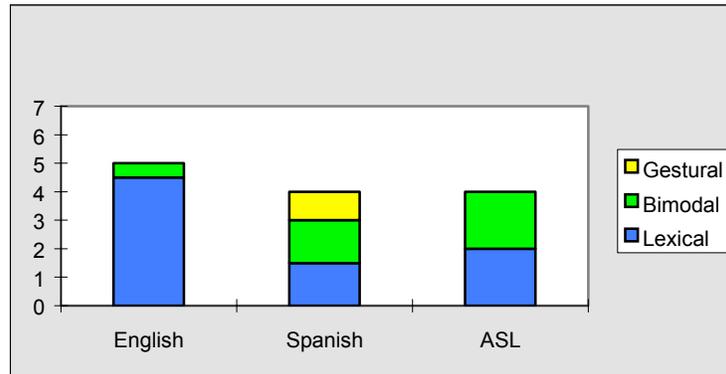
**Table 3.** Path conceptual items

| Language | Lexical Items | Bimodal Items | Gestural Items |
|----------|---------------|---------------|----------------|
| English  | 1.5           | 3.5           | 1.5            |
| Spanish  | 1.5           | 1.5           | 3              |
| ASL      | 0             | 1.5           | 4              |

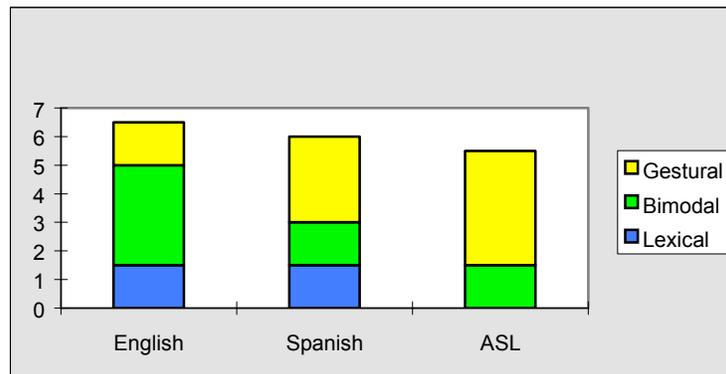
**Table 4.** Manner conceptual items

| Language | Lexical Items | Bimodal Items | Gestural Items |
|----------|---------------|---------------|----------------|
| English  | 4             | 1             | 0              |
| Spanish  | 2.5           | 1.5           | 1              |
| ASL      | 0             | 1             | 2              |

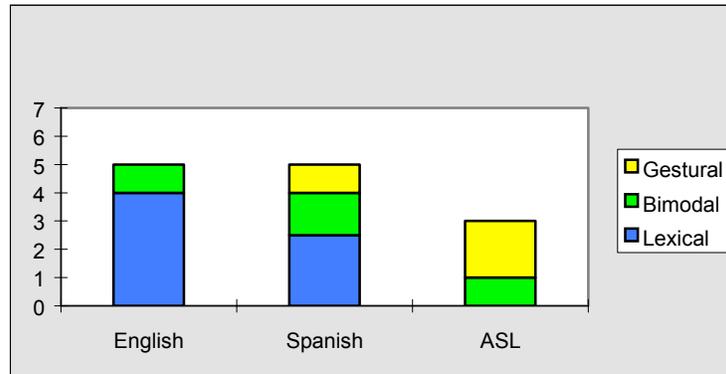
**Figure 2.** Figure/Ground conceptual items



**Figure 3.** Path conceptual items



**Figure 4.** Manner conceptual items



To summarize, Figure/Ground information is expressed lexically in all languages, with slight supplementation from gesture in Spanish. Path information is expressed gesturally in all languages, with some supplementation from lexical items in English and Spanish. Manner information varies considerably: English expresses manner lexically, ASL expresses it gesturally, and Spanish seems to employ a mixed strategy.

### 3. Conclusions

The data suggest a significant role for gesture in equalizing the expression of information across languages and modalities. Let us go through the hypotheses one by one.

#### A. Similar amounts of conceptual information

This hypothesis is basically supported by the data. Subjects produced approximately the same number of items overall. If gesture were not considered, Spanish subjects would have produced significantly less information than ASL subjects and English subjects. English subjects also expressed additional items through gesture. Thus, gesture may be said to play a role in equalizing the amount of information expressed.

Overall, the data suggest that English conveys more information through lexical means, ASL conveys more information through mapped items, and Spanish conveys equal amounts through both modes.

- B. i. Similar types of information in speech and lexical sign elements
- ii. Similar types of information in gesture and gestural sign elements

These hypotheses are partially supported. We seem to find that certain types of conceptual items are preferentially expressed through one mode or the other: Figure/Ground information always shows up in lexical elements, and Path infor-

mation always shows up in gestural elements. Languages vary as to how much these types show up in their non-preferred mode. Thus, for example, all subjects support Figure/Ground information with some gestural elements, but this is more prevalent for ASL and Spanish. Similarly, ASL only rarely supports Path information with lexical elements, while English and Spanish do this to a considerable degree, adding non-gestured Path items in their speech. These two tendencies make a great deal of sense. Identification of things (Figure/Ground elements) is most easily accomplished through memorized lexical elements or category names, while the details of Paths are more easily shown through free gesture than through fixed lexical categories.

On the other hand, Manner shows no overall mode preference. English subjects express it largely lexically, adding no new information through gesture; ASL subjects show the opposite pattern; and Spanish subjects present new information through both modes, though there may be a trend toward lexical expression. This may be related to differences in English and Spanish lexicalization patterns, as discussed below.

#### C. Similar rates of information presentation

This hypothesis is not supported. We see wide variation in rate of presentation, with English the slowest, ASL faster, and Spanish the fastest. This result may be an artifact of the pilot study conditions, in which Spanish speakers watched seven minutes of videotape before telling the stories, while users of the other languages watched one minute or less at a time. The larger study will have uniform conditions for all subjects.

#### D. Spanish uses gesture to add more Path and Manner information.

This hypothesis is supported, strongly for Path items and more weakly for Manner items. Were gesture not considered, Spanish speakers would be judged to express less Path information than either other language; with gesture, they express roughly the same amount. Some Manner information is also added by gesture, but this effect may not be significant.

### 4. Summary

We find that in both of the spoken languages, subjects express a significant amount of additional information through gesture; this amount is greater for the verb-framed language (Spanish) and seems to offset the constraints of the verb-framed language type. The contribution of gesture produces a rough equalization of amount of information expressed. As mentioned above, there is no consensus among researchers on the function of co-speech gestures. They may communicate information, help the speaker think about a topic, facilitate access to a lexical item, or some combination of these. Our results show that gestures do not simply

## *Comparing Speech/Gesture and Sign Language*

reinforce the information presented through speech; rather, new information is presented through the gestural modality.

We note as well that particular categories of conceptual items appear universally in speech or universally in gesture, optionally supported by the other mode. This pattern would not have been apparent had we not separated gestural and lexical elements within sign language, and brought in gesture as well as speech for cross-modal comparisons. We conclude that the methodology of comparing speech/gesture complexes with each other and with sign languages produces interesting results, and we recommend it as the standard for future typological investigations.

### **References**

- Galvan, Dennis, and Sarah F. Taub. Forthcoming. The Encoding of Motion Information in American Sign Language. In R. Berman, D. Slobin, S. Strömquist, and L. Verhoeven (eds.), *The Frog Story Revisited*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fauconnier, Gilles, and Mark Turner. 1996. Blending as a Central Process of Grammar. In Adele Goldberg (ed.), *Conceptual Structure, Discourse, and Language*. Stanford, CA: CSLI, 113-130.
- Slobin, Dan I. 1996. Two Ways to Travel: Verbs of Motion in English and Spanish. In M. Shibatani and S. A. Thompson (eds.), *Essays in Semantics*. Oxford: Oxford University Press.
- Kendon, Adam. 1986. *Some Reasons for Studying Gesture*. Amsterdam: Mouton de Gruyter.
- Liddell, Scott. 2000. Blended Spaces and Deixis in Sign Language Discourse. In D. McNeill (ed.), *Language and Gesture*. Cambridge: Cambridge University Press.
- McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- Talmy, Leonard. 1985. Lexicalization Patterns: Semantic Structure in Lexical Forms. In T. Shopen (ed.), *Language Typology and Syntactic Description, Vol. III: Grammatical Categories and the Lexicon*. Cambridge: Cambridge University Press.

Department of Linguistics  
Gallaudet University  
Dawes House  
800 Florida Avenue, NE  
Washington, DC 20002

Sarah.Taub@gallaudet.edu

Department of Foreign Languages  
Gallaudet University  
Hall Memorial Building  
800 Florida Avenue, NE  
Washington, DC 20002

Pilar.Pinar@gallaudet.edu