Berkeley Linguistics Society

Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society: Special Session on Caucasian, Dravidian, and Turkic Linguistics (2000)

PROCEEDINGS OF THE TWENTY-FIFTH ANNUAL MEETING

OF THE

BERKELEY LINGUISTICS SOCIETY

February 12-15, 1999

SPECIAL SESSION

ON

CAUCASIAN, DRAVIDIAN, AND TURKIC LINGUISTICS

edited by

Jeff Good

Alan C.L. Yu

Berkeley Linguistics Society

PROCEEDINGS OF THE TWENTY-FIFTH ANNUAL MEETING

OF THE

BERKELEY LINGUISTICS SOCIETY

February 12-15, 1999

SPECIAL SESSION

ON

CAUCASIAN, DRAVIDIAN, AND TURKIC LINGUISTICS

edited by

Jeff Good

Alan C.L. Yu

TABLE OF CONTENTS

SPECIAL SESSION:
CAUCASIAN, DRAVIDIAN, AND TURKIC LINGUISTICS

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

Interpreting genitives in Turkish
       MÜRVET ENÇ

Morphosemantics of deverbal adjectives in Malayalam
       K.P. MOHANAN

GENERAL SESSION
AND
PARASESSION ON LOAN WORD PHENOMENA
FOUND IN BLS 25

PARASESSION: LOAN WORD PHENOMENA

♦   ♦   ♦

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

ACD, QR, and frozen scope
    BENJAMIN BRUENING

A model for the construction of common ground in interpreted discourse
    BRAD DAVIDSON

Loan word phonology in Optimality Theory
    JUNKO ITÔ and ARMIN MESTER

A local treatment of nonlocal relativization: A constructional approach
    JONG-BOK KIM

The origins and development of Chinese classifiers: A grammaticization perspective
    FENGXIANG LI

Construction but no constructions: Doing without the lexicon
    ALEC MARANTZ

AUTHOR INDEX

PREFACE

We present to you the proceedings of the Special Session of the 25th Annual
Meeting   of the Berkeley Linguistics Society. The topic   was Caucasion,
Dravidian, and Turkic Linguistics, and the conference was held on February 12th,
1999. We would like to thank all the past BLS conference organizers for their
helpful advice, and we would also like to thank our co-organizers: Steve Chang,
Lily Liaw, and Josef Ruppenhoffer. We would also be remiss if we neglected to
thank all of the volunteers who made sure the conference was a success. Finally,
we would like to thank the Berkeley Linguistics Department for its administrative
assistance. We hope you enjoy the volume.

BLS 25S Editors
Jeff Good
Alan C. L. Yu

The historical geography of pharyngeals and laterals in the Caucasus

JOHANNA NICHOLS
University of California, Berkeley

1. Introduction

The signature "exotic" sounds that help characterize the Caucasus as a linguistic area are pharyngeal series of consonant phonemes, pharyngealization as a feature of syllables or even morphemes, and lateral obstruent series. None of these signature sound classes, however, is pan-Caucasian in its distribution. Here it will be shown that these sounds have similar geographies that span two of the indigenous language families and bring the frontier branches of one indigenous family into the structural fold of its neighbor. All are absent from the third indigenous family as well as later entrants. This paper will trace the chronology, geography, and structural phonology of these consonant series and propose that pharyngeal and lateral segments are ancient endemic features of the Caucasus, reconstructible to a geographical area but not necessarily to any family or families.

The Caucasus is known for its large number of languages and language families. There are three indigenous language families, i.e. families found only in the Caucasus and with no outside kin; basic information for them is given in (1). Then there are languages which do have outside kin and have entered most recently: Ossetic, an Iranian language related to or descended from Scythian of the Iron Age steppe and resident in the Caucasus some 3000 years or more; Tat and Talysh, Iranian languages of the same branch as Persian and resident in the Caucasus for perhaps a millennium; and Karachay-Balkar, Kumyk, and Azeri, languages of the Turkic family resident in the Caucasus since the early middle ages. These are the languages spoken in or near the Great Caucasus range; others, notably Armenian, are spoken in the Little Caucasus to the south, outside the purview of this paper.

(1) The indigenous language families of the Caucasus, their ages, and numbers of daughter languages.

Nakh-Daghestanian (Northeast Caucasian) (age ca. 6000+) (~30 langs.)
    Nakh: Chechen, Ingush, Batsbi
    Daghestanian: Avar-Andic and Tsezic (several languages each)
              Lak
              Dargic
              Lezghian (several languages)
Northwest Caucasian (Abkhaz-Circassian) (age unknown; deep) (3 langs.)
    Abkhaz-Abaza
    † Ubykh
    Circassian: Adyghe, Kabardian
Kartvelian (South Caucasian)   (age ca. 4000 - 4500) (4 langs.)
    Svan
    Georgian-Zan: Laz; Mingrelian, Georgian

2. Pharyngeals and pharyngealization

Most Nakh-Daghestanian languages have a contrast between plain and pharyngealized syllables. Pharyngealization is a phonation type whose occurrence

is favored or disfavored in individual languages by syllable-initial consonants (e.g., favored by uvulars in many Daghestanian languages, absolutely incompatible with them in Nakh) and whose inheritance is therefore not always straightforward. The phonetics of pharyngealization varies from language to language; Kodzasov 1990 classifies it into two types which will be called r-colored and 'ain-colored here. 'Ain-colored pharyngealization is basically epiglottalization. For the r-colored type the constriction is higher than the epiglottis and entails some tensing around the root and lower body of the tongue, and the acoustic effect is often similar to that of American English postvocalic /r/.

In addition to pharyngealization, a number of languages have phonemic pharyngeal segments (which are phonetically epiglottal): /ʕ/ = [ʔ] and /ħ/. These are often structurally equivalent to /ʔ/ or /h/ respectively in a pharyngealized syllable. /ʕ/ is marginal in many of the languages, often limited to baby talk, onomatopoeia, and Arabic loans.

For the phonetics and phonology of pharyngeals and pharyngealization in the various languages see Kodzasov 1990, chapters in Smeets ed. 1994, Catford 1983, Ladefoged and Maddieson 1996.

It is sometimes useful to refer jointly to pharyngeal segments and pharyngealization. Here the term *pharyngeal component* will be used generically for the two.

2.1. Nakh-Daghestanian

(2)-(4) give three cognate sets for Nakh-Daghestanian showing pharyngeals and pharyngealization.[1, 2] Nakh-Daghestanian words generally display regular reflexes for postvocalic consonants, but initial consonants are highly variable; this is due in part to ancient prefixation and in part to contamination by the noun's gender marker (prefixed to verbs in agreement) (a clear example is (2)). Nakh-Daghestanian languages have up to six gender classes, and they are retained in a regular hierarchy: 2 + 3 < 1 < 4 < 5. In reconstructing genders, the firmest evidence therefore comes from the languages which retain the most genders. Gender 3 in a three-gender language could reflect 3, 4, or 5, but gender 3 in a five-gender language points more clearly to gender 3. (There have of course been changes in the gender of individual words.) All three of the words in (2)-(4) were probably of Proto-Nakh-Daghestanian gender 3. The genders of the daughter forms, and the total number of genders in each language, are given in (2)-(4) to reassure wary readers that there is more than just a postvocalic consonant to reconstruct. (For Nakh-Daghestanian gender see Nichols 1990, 1993, Schulze, in preparation.)

In the word for 'eye', in (2), there is good evidence for reconstructing pharyngealization. A pharyngeal component is found in most branches; it is found in nearly every cognate that is unprefixed and retains the PNEC first syllable (i.e. not Archi), in those daughter languages that have pharyngeal or epiglottal segments. It is preserved in Nakh, together with prefix, as pharyngealization. It is lost in Udi, where the prefixal element is preserved, though in principle Udi should be able to have pharyngealization on this word. Agul has pharyngeal segments in most dialects but has lost pharyngealization in this word.

The words for 'apple', in (3), point to an initial *h. Many of the daughter languages lose initial *h and/or *ħ, but the evidence for both pharyngeal quality and [h]-like phonetics is clear in this set.

The words for 'boar', in (4), show evidence of a pharyngeal segment in Nakh, pharyngealization in Tsez, Archi, and Udi. There is a labial initial -- evidently a prefix -- and gender 3 (whose marker is a labial) in all Daghestanian branches, and

a different prefix -- a pharyngeal segment -- and gender 5 in Nakh. The pharyngeal component evidently belonged to the root element and not the prefix.

These examples are typical in that they give good evidence for an ancestral pharyngeal component without displaying perfectly regular correspondences for that pharyngeal component. The pharyngeal component can be reconstructed as a lexical property of individual words and as a phonological element in the protolanguage. Usually, as in (2) and (4), pharyngealization seems to be a more plausible reconstruction than a pharyngeal segment, though occasionally, as in (3), there is evidence for a segment. So far I have found no cognates that suggest a Proto-Nakh-Daghestanian *ʕ.

(2) # (b)ʕul / # (b)ʕal (3) 'eye'. #b, where present, is due to gender contamination (the prefix for gender 3 is *b-). Daghestanian forms from Kibrik & Kodzasov 1990:10-11 (#3).

		Gender
Nakh:		
Chechen	bʕar, bʕär-jg	5, 3 / 5
Ingush	bʕar, bʕar-jg	3 / 5
Batsbi	bʕar, bʕar-k'	3 / 5
Avar-Andic-Tsezic:		
Avar (Chadakolob)	ber	3 / 3
Andic	(no cognates)	
Hunzib	hare	4 / 5
Bezhta: Tljadal	ħäre	4 / 5
Xoshar-Xota	häre	4 / 5
Lak	(no cognate)	
Dargva: Chirag	ʕule	3 / 3
Lezghian:		
Lezghi: Akhtyn	ul	
Tabassaran (N, S)	ul	2 / 2
Agul: Burschag	il	
others	ul	
Rutul	ul	3 / 4
Tsaxur	uly	3 / 4
Kryts	ʕül	3 / 4
Budux	ʕül	3 / 4
Archi	lur	3 / 4
Udi	pul	
Xinalug	pil	3 / 4

(3) Three variant forms for 'apple' (all probably gender 3). Nakh glosses: 'p' = 'plum', 'm' = 'medlar', 'a' = 'apple'. All others mean 'apple'. ==: form could as well be in this column. The third form, found only in Nakh, is evidently a borrowing into Proto-Nakh from an unattested extinct language or branch. (It is the only one of the three that refers unambiguously to a domesticate, and is a plausible cultural borrowing.) Daghestanian forms from Kibrik & Kodzasov 1990:104-5 (#215).

	# ħam(V)c	# ħam(V)č	* ʕaž (#ʕamč' ?)
Nakh:			
Chechen	ħamc 'm' 3 /5	ħač, ħieč- 'p' 3 /5	ʕaž 'a' 3 /5
		eel+ħamch 'hawthorn' 4 /5	
Ingush	ħamasjk 'm' 5 /5	ħač, ħač- 'p' 3 /5	ʕaž 'a' 3 /5
	< *ħamVc-ig		
Avar-Andic-Tsezic:			
Avar		ʕič, ʕoč-o- 3 /3	
Andi		inči #3 /5	
Akhvakh	==	ʕeče 3 /3	
Chamalal		eš 3 /3	
Chamalal: Gigatli	== ?	míči ‖ mísči 3 /3	
Tindi		eči 3 /3	
Tsez		heneš 3 /4	
Xvarshi		ẽš 3 /5	
Ginux		iši ‖ ʕiši 3 /4	
Bezhta		ẽš 3 /5	
Lak		hičw, hič-u- 3 /4	
Dargva	ʕinc 3 /3		
Lezghian:			
Lezghi		ič	
N. Tabassaran		waʕč 2 /2	
Agul (Richa, Burkixan)		ħač	
Rutul		eč 3 /4	
Tsaxur		eč 3 /4	
Kryts		ječ 3 /4	
Budux		ječ 3 /4 .	
Archi		aʕnš 4 /4	
Udi		ešʕ	
Xinalug		məč 4 /4	

(4) * ...ƛ... / * ...ƛƛ'... in 'wild boar, pig' (3). Gloss is 'pig' in Nakh, 'boar, wild boar' elsewhere. The Akhvakh and Xvarshi forms must be borrowed from other NEC languages (both from Andic sources). The Avar-Andic protoform must have been borrowed from Tsezic. Daghestanian forms: Kibrik and Kodzasov 1990:77 (#148). Batsbi: Aliroev 1970:89.

	# (bV) ƛo		# (bV) ƛƛ'o		Irregular reflexes
Nakh:					
Chechen	ħaqa, pl. hʕaqar-	5 /5			
Ingush	ħaqa, pl. hʕaqar-	5 /5			
Batsbi	qa, pl. qar-	3 /5			
Avar-Andic-Tsezic:					
Avar Ch.	botón	3 /3			
Akhvakh					boɬõ 3 /3
Chamalal	boɬũ	3 /5			
Tindi	boɬon	3 /5			
Tsez			baʕƛo	3 /4	
Xvarshi					bolhon (!) 3/5
Ginux			boƛi	3 /4	
Hunzib			buƛu	3 /5	
Bezhta			buƛo	3 /5	
Lak			burk'	3 /4	
Dargva	(no cognate)				
Lezghian:					
Lezghi			wak		
Agul			wakk		
Tsaxur			wok	3 /4	
Kryts			vak	3 /4	
Budux			vək	3 /4	
Archi			boʕƛ'	3 /4	
Udi			boq'ʕ		

2.2. Northwest Caucasian

Proto-Circassian had *ħ but no *ʕ (Kuipers 1975). Abkhaz and Ubykh also have /ħ/ but no /ʕ/. Kabardian has marginal /ʕ/, found only in baby words, onomatopoeia, and Arabic loans (in the latter it can be replaced by /ħ/). Only Abaza has an /ʕ/. No language in the family has a syllabic feature of pharyngealization. As far as can be determined from available sources, the quality of pharyngeals is always epiglottal in this family.

2.3. Kartvelian

No member of the Kartvelian family has pharyngeals or pharyngealization.

2.4. Discussion

(5) is a table showing the numbers of languages in the eastern and western branches of Nakh-Daghestanian that have pharyngeal components of various kinds.

(Note that 'eastern' and 'western' are purely geographical terms; this is not a genetic subgrouping.) There is a geographical skewing to the distribution of pharyngeals and pharyngealization. We have seen that pharyngealization is more securely reconstructible and more widely attested than pharyngeal segments in Nakh-Daghestanian, while Northwest Caucasian languages have pharyngeal segments but no syllabic pharyngealization. Within Nakh-Daghestanian, languages with pharyngeal segments but no pharyngealization are found predominantly in the western languages, which thereby resemble the Northwest Caucasian language family to the west. Pharyngealization is more common in the eastern Nakh-Daghestanian languages than in the western ones, and *r*-colored pharyngealization is almost exclusively eastern; both phenomena are entirely lacking in Northwest Caucasian. Thus the western Nakh-Daghestanian branches resemble Northwest Caucasian. There is an east-to-west clinal distribution which transgresses family boundaries.

(5) Distribution of pharyngeal segments and syllabic property of pharyngealization in Nakh-Daghestanian languages. Entries are numbers of languges. W = western branches (Nakh, Avar-Andic-Tsez), E = eastern branches (Lak, Dargi, Lezghian). NWC = features also found in Northwest Caucasian. (In all of these counts, marginal attestations were counted as non-attestations. The same geographical distribution was found when they were counted as attestations.)

	E	W	Total	Geography	NWC?
segments only	3	6	9	western	yes
syllabic feature only	4	0	4	eastern	no
both	8	6	14		no
neither	1	1	2		
segments only or both	11	12	23	(western)	yes
syllabic feature ,"ain"-colored	7	5	12		no
syllabic feature , "r"-colored	12	1	13	eastern	no
syllabic feature (total)	13	6	19	eastern	no
"ain" quality	12	13	25		yes
"r" quality	12	1	13	eastern	no
Total no. languages	17	14	31		

Now, pharyngeal component phenomena are extremely resistant to change in the branches of Nakh-Daghestanian and probably also in Northwest Caucasian. Furthermore, they evidently do not spread areally with any readiness: Kartvelian languages and Ossetic, after some millennia of residence in the Caucasus, have neither pharyngeals nor pharyngealization, though they have other phonological features of the Caucasus including ejectives and uvulars. Given this resistance to spread, how could a cline which crosscuts family boundaries have taken shape?

Northwest Caucasian and Nakh-Daghestanian are not now in contact; they are separated by Kartvelian, Ossetic, and Karachay-Balkar. Suppose they represent the remnants of a once-continuous linguistic area later split by the spread of Kartvelian (by c. 4500 B.P.) and the entries of Ossetic (c. 3000-4000 B.P.?) and Karachay-Balkar (ca. 1000-700 B.P.). Then, prior to these movements, Northwest Caucasian or languages of its type must have extended farther east, and/or Nakh-Daghestanian may have extended farther west, and/or there may have been some now-extinct third family in the central Caucasus.

Whatever languages anciently lay between Northwest Caucasian and Nkah-Daghestanian must have had pharyngeals and perhaps also pharyngealization. Areality and the east-west cline in pharyngealization features must have come about not as a result of borrowing of pharyngeals or pharyngealization, but as a result of fine-tuning one's pharyngeal and/or pharyngealization system to bring it closer to a neighbor's system. Note that easternmost Northwest Caucasian and westernmost Nakh-Daghestanian have *not* lost pharyngeals or pharyngealization under the influence of neighbors; this suggests that the vanished intermediary did not lack pharyngeals.

The pharyngeal cline could have taken shape only slowly and gradually over a very long period of time. To judge from what happens in the modern Nakh-Daghestanian languages, pharyngeal segments can arise in three ways. In many of the languages, initial vowels are preceded by an automatic glottal stop, and if the word is pharyngealized then that glottal stop will be pharyngealized, i.e. will be a pharyngeal or epiglottal stop. If this segment is phonemicized, the result is a pharyngeal segment. The fact that in many languages pharyngeal segments occur only in onomatopoetic words, baby words, and Arabic loans indicates that pharyngeal segments can arise as a result of borrowing or incorporation into the standard register. To judge from the Daghestanian languages, the presence of syllabic pharyngealization seems to prime the language for innovating or borrowing pharyngeal segments, if they were not there before. (The Turkic, Iranian, and Kartvelian languages of the Caucasus do not acquire pharyngeals as a result of Arabic borrowings.) Finally, uvulars may turn into pharyngeals, as happens in Avar and Nakh; an example is in §3 below. There is no evidence at all on how syllabic pharyngealization may arise; in the Caucasus it occurs only in Nakh-Daghestanian, where it is inherited from the protolanguage.

3. Laterals.

Proto-Nakh-Daghestanian had a full series of lateral affricates (plain, fortis, ejective, fortis ejective) and a voiceless fricative. Correspondences for some of these are shown in (6). The affricates of Proto-Nakh-Daghestanian are shown in (7). Only affricates distinguish plain from fortis ejectives, and on this evidence as well as on the phonetic witness of the daughter languages the uvular obstruents are classified as affricates, not stops. (8)-(9) give illustrative cognate sets for the fortis ejective. (For more cognates see Nichols 1993.)

Laterals have merged with uvulars in some languages (Proto-Nakh, Rutul, Udi) with velars in others (Lak, Dargi, most Lezghian languages), and with dentals in Chadakolob Avar. They are preserved as laterals now only in the Avar-Andic-Tsezic branch and in Archi, which belongs to the Lezghian branch but is geographically separated from its sisters and influenced by Avar and Lak. Proto-Nakh had a voiceless lateral fricative (preserved only in Batsbi) of uncertain origin; it does not descend from a Proto-Nakh-Daghestanian voiceless lateral fricative. Phonetically, the Daghestanian laterals have variously palatovelar and palatoalveolar articulation of the stop component (see Kodzasov 1990); the release is either lateral or (as Maddieson 1999 shows for Archi) frontal (through a central tongue groove). In view of its reflexes and phonetics, the Proto-Nakh-Daghestanian lateral series

appears to have actually been a velar or palatovelar (or perhaps simply palatalized) counterpart to the uvular series. True phonetic lateral pronunciation is a secondary development of Avar-Andic-Tsezic, and a secondary lateral obstruent develops in Nakh. That is, the westernmost branches have laterals.

Northwest Caucasian has lateral fricatives in all three branches, and an affricate in Ubykh (and marginally in Abaza). These are phonologically and phonetically laterals and show little tendency to merge with any other series. Neither Kartvelian nor any of the non-indigenous languages have lateral obstruents.

It appears, then, that the phonetic realization of the Proto-Nakh-Daghestanian "lateral" series as actual laterals in Avar-Andic-Tsezic is a secondary development which brings that branch into closer phonetic and phonological alignment with Northwest Caucasian to the west; and likewise for the secondary lateral fricative of Proto-Nakh. That is, lateral obstruents occur in the west of the Caucasus: in Northwest Caucasian (where they are evidently native) and in westernmost Nakh-Daghestanian (where the realization as laterals is secondary).

(6) Regular correspondences for selected laterals (the best-attested ones) in Northeast Caucasian.

	*ʎʎ'	*ʎ'	*ɬɬ	Articulation
Nakh (all languages)	ʕ	q'	x	Uvular, pharyngeal
Avar-Andic-Tsezic:				
Chadakolob Avar	ʎ'	t'		Lateral, dental
Standard Avar	ʎ' = [ʎʎ']	ʎ'	ɬ	Lateral
Andic	ʎ' ~ [ʎʎ']	ʎ' (?)	ɬ	Lateral
Tsezic	ʎ'	ʎ	ɬ	Lateral
Lak	k'	k'	xx´/ šš	Velar
Dargva: Chirag	g	k'	x´/ š ?	Velar
Lezghian				
Akhtyn Lezghi	k'	?	?	Velar
Standard Lezghi	k	q'	ɣ́	Velar, uvular
Tabassaran	kk	k'	x´ / š	Velar
Agul	kk	k'	x´, š	Velar
Rutul	q'	q'	x´/ ħ	Uvular; velar/phar.
Tsaxur	k'	k'	x´	Velar
Kryts	k'	k' (q'?)	x´	Velar (& uvular?)
Budux	k'	(q'?)	x´	Velar (& uvular?)
Archi	ʎ'	k'	ɬ	Lateral
Udi	q	Ø / q'	x	Uvular
Xinalug	k	k'	x´	Velar

(7) Proto-Nakh-Daghestanian voiceless stops and affricates

p	t	c	č	ƛ	k	q
		cc	čč	ƛ		qq
(p')	t'	c'	č'	ƛ'	k'	q'
		cc'	čč'	ƛƛ'		qq'

(8) #ƛƛ'Vn-i 'winter'. Oblique stems are cited for *CV nominatives to show the unity of the root. (Daghestanian forms: Kibrik and Kodzasov 1990:210 (#451).)

		Nominative	Oblique
Nakh:			
	Chechen	ʕa	ʕän-
	Ingush	ʕa	ʕan
	Batsbi	ʕa	
Avar-Andic-Tsezic:			
	Chadakolob Avar	ƛ'in	
	Andi		ƛ'ina (adverb)
	Tsez	eƛni	
	Xvarshi (Inxokvari)	yƛnu	
	Hunzib	ƛini	
	Bezhta	ƛine	
Lak		k'wi	k'wi-nti-l
Dargi: Chirag		ga	gane (adverb)
Lezghian: Rutul			q•ynni (?) (adverb)

(9) # (Da-)aƛƛ'e-R 'bone'   (4 or 5). Daghestanian forms: Kibrik and Kodzasov 1990:38-9 (#66).

			Gender
Nakh:			
	Chechen	däʕaxk	5 / 5
	Ingush	t'iexk	5 / 5
	Batsbi	dʕeʔ	4 / 5
	Proto-Nakh	*t'eʕa-xk or *t'aʕe-xk (or *deʕa-xk, *daʕe-xk)	
Avar-Andic-Tsezic:			
	Chadakolob Avar	raƛ'a	3 / 3
	Andi	roƛ'i	#3 / 5
	Akhvakh	raƛƛ'i	3 / 3
	Chamalal: L. Gakvari	jaƛ'il	4 / 5
	Gigatli	raƛ'il	4 / 5
	Tindi	raƛ'i	3 / 3
	Tsez	ƛuza	3 / 4
	Xvarshi (Inxokvari)	ƛozol	3 / 5
	Ginux	ƛužej	4 / 4
	Bezhta: Tljadal	ƛöra	4 / 5
	Xoshar-Xota	ƛõra	4 / 5
Lak		ttark', tturč'-	4 / 4
Dargi: Chirag		derg	3 / 3
Lezghian:			
	Akhtyn Lezghi	k'rab ±	
	and also:	kwarg	
	N. Tabassaran jirkki		2 / 2
	Agul (Burschag)	jirkk	
	other Agul	irkk	
	Rutul	q'yryˋb ±	4 / 4
	Tsaxur	barkw'	3 / 4
	Kryts	k'ärap' ±	4 / 4
	Budux	k'erep' ±	3 / 4
	Udi	uq'ʕen	
	Xinalug	ink'	4 / 4

± Words marked with this symbol may be connected to Armenian *k'arapn* (W. Schulze, p.c.). Within Lezghian, though, they show the regular correspondence for *ƛƛ'.

4. Conclusions.

We have seen that the rare sound classes that are endemic to the Caucasus -- pharyngeal segments and lateral obstruents -- are not Caucasus-wide or even North-Caucasus-wide but have an east-west clinal distribution that crosscuts family lines. Both are anciently indigenous to the Caucasus, secondarily acquired by Nakh-Daghestanian, found now chiefly or only in the central and western part of the north Caucasus, and native to Northwest Caucasian. Proto-Nakh-Daghestanian must have had a syllabic feature of pharyngealization, and to judge from the geographical and phylogenetic distribution of *r*-colored and *'ain*-colored pharyngealization in the modern languages the ancestral phonation type is most likely to have been *r*-

colored. Epiglottalization or *'ain*-colored pharyngealization appears to be a secondary development, favored in the westernmost branches of the family. There may or may not have been a pharyngeal fricative *ħ in the protolanguage; there·was no *ʕ, although the western branches now have it as a well-established phoneme. Proto-Nakh-Daghestanian also seems to have had a palatovelar affricate series that was reanalyzed as lateral in the westernmost Daghestanian branch. All of these developments appear to be accommodations to the consonantism of the Northwest Caucasian languages, which have epiglottal and lateral segments.

The historical scenario that could explain this distribution of pharyngeals and laterals is consistent with what is known about the linguistic prehistory of the Caucasus. Proto-Nakh-Daghestanian entered the Caucasus from the southeast, dispersing in the vicinity of today's Azerbaijan at least about 6000 years ago. (See Schulze 1999:171ff., Nichols 1999 for the breakup and dispersal.) As the Avar-Andic-Tsezic and Nakh branches spread west and north to their present locations, they evidently came into contact with languages that had epiglottal segments and lateral obstruents, and acquired these sounds by phonetic adjustment of existing phoneme series and through sound changes that created voiced pharyngeal stops from uvulars. The epiglottal quality of the new sounds also affects the inherited pharyngeal phonation type.

Northwest Caucasian is long indigenous to the Caucasus. The lateral and pharyngeal inventories of its daughter languages do not appear to have been affected at all by the sound systems of neighboring languages. Of their neighbors their phonological type has affected only westernmost Nakh-Daghestanian.

Proto-Kartvelian dispersed around 4000-4500 years ago and spread gradually to its present location with Georgian adjacent to and partly intermingled with Abkhaz (Northwest Caucasian). The contact of Kartvelian with the Northwest Caucasian phonological type may have been less in duration and/or intensity than that of westernmost Nakh-Daghestanian, not sufficient to permit the spread of pharyngeal or lateral series into Kartvelian. The same is true of Ossetic - Northwest Caucasian contacts.

The present distribution of pharyngeals and laterals results from westernmost Nakh-Daghestanian coming into the sphere of influence of the Northwest Caucasian phonological type. Now, there has been no known historical contact between Avar or Andic or Tsezic and any Northwest Caucasian language; and contact between Nakh and Kabardian is believed to have begun in the late middle ages when Ingush and western Chechen outposts in the north central Caucasus foothills grew into important settlements. The known history cannot account for the development of epiglottals and laterals in Proto-Avar-Andic-Tsezic and Proto-Nakh. Rather, it must be that the relevant contacts accompanied the spread of these protolanguages to their protohomelands in the north central Caucasus. No living Northwest Caucasian language has ever been spoken in the central Caucasus, on either the north or the south slope, but in the late Bronze and early Iron Ages the structural type, or sphere of influence, to which Northwest Caucasian belongs must have extended to the central Caucasus. The influence could have involved substratum, direct borrowing, or less direct diffusion; we will probably never know. All that can be said is that westernmost Nakh-Daghestanian came into contact with the eastern province of the Northwest Caucasian type and acquired certain striking traits as a result, while Kartvelian, Ossetic, and Turkic languages have not acquired these traits and may not have had the ancient contact. The result, millennia later, is that two exotic phonological features of the western Caucasus now extend, discontinuously, across family boundaries and show up in western Nakh-Daghestanian.

The linguistic geography and historical phonology of pharyngeals and laterals make it possible to reconstruct an ancient language area -- the western and central Caucasus -- independent of any particular language or family now present in the Caucasus. This area has a geography and a structural type (and it has a small lexical stock, not covered here), as well as a sociolinguistics that has enabled it to draw in languge families over time; but it has no phylogenetic identity -- no identifiable member families -- at all.

Notes

[1] Nikolayev and Starostin 1994 offer many putative cognate sets and reconstructions for Nakh-Daghestanian and its branches, as part of a reconstructin of a putative North Caucasian comprising Northwest Caucasian and Nakh-Daghestanian. These two families have not in fact been shown to be related. Nikolayev and Starostin proceed by assuming relatedness and then assembling cognate sets so as to maximize recurrent correspondences. This entails splitting up a number of actual Nakh-Daghestanian cognate sets and distributing the cognates among two or more putative North Caucasian cognate sets. The cognates and correspondences offered here are those demanded by Nakh-Daghestanian data. They are described in Nichols 1993, and sets reproduced here come from my own database of Nakh-Daghestanian cognates.

[2] In these cognate sets the ancestral forms are not true reconstructions but quasi-reconstructions pending fuller analysis of the morphology. Hence they are marked with # rather than an asterisk. An asterisk is used for branch protoforms which are true reconstructions. In protoforms and modern language forms, both pharyngealization and the voiced pharyngeal segment /ʕ/ are transcribed with the same symbol. It represents pharyngealization if adjacent to a consonant, the pharyngeal segment otherwise.

References and sources

Catford, J.C. 1983. Pharyngeal and laryngeal sounds in Caucasian languages. 345-50 in Diane M. Bless and James H. Abbs, eds., *Vocal Fold Physiology: Contemporary Research and Clinical Issues.* San Diego: College-Hill Press.

Charachidze, Georges. 1989. Ubykh. 357-459 in B. George Hewitt, ed., *The Indigenous Languages of the Caucasus, vol. 2: The Northwest Caucasian Languages.* Delmar, NY: Caravan Books.

Colarusso, John. 1992. *A Grammar of the Kabardian Language.* Calgary: University of Calgary Press.

Comrie, Bernard. 1981. *The Languages of the Soviet Union.* Cambridge: Cambridge University Press.

Giginejshvili, B. K. 1977. *Sravnitel'naja fonetika dagestanskix jazykov.* Tbilisi: Tbilisi University Press.

Kibrik, A. E., and S. V. Kodzasov. 1988. *Sopostavitel'noe izuchenie dagestanskix jazykov: Glagol.* Moscow: Moscow University Press.

----. 1990. *Sopostavitel'noe izuchenie dagestanskix jazykov: Imja. Fonetika.* Moscow: Moscow University Press.

Kodzasov, S. V. 1990. Fonetika [dagestanskix jazykov]. Kibrik and Kodzasov 1990:311-47.

Kuipers, A. H. 1975. *A Dictionary of Proto-Circassian Roots.* Lisse: Peter de Ridder Press.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell.

Nichols, Johanna. 1993. The Nakh-Daghestanian consonant correspondences. Presented at the Conference on Non-Slavic Languages of the [Former] USSR, University of Chicago.

----. 1990. The Nakh evidence for the history of gender in Nakh-Daghestanian. H. I. Aronson, ed., *The Non-Slavic Languages of the USSR: Linguistic Studies,* 158-75. Chicago: Chicago Linguistic Society (University of Chicago).

----. 1999. The origin of the Chechen and Ingush. *Annual of the Society for the Study of Caucasia*, to appear.

Nikolayev, S. L., and S. A. Starostin. 1994. *A North Caucasian Etymological Dictionary.* Moscow: Asterisk.

Schulze, Wolfgang. 1999. *Person, Klasse, Kongruenz: Fragmente einer Kategorialtypologie des einfache Satzes in den ostkaukasischen Sprachen,* Vol. 1: Die Grundlagen. Munich: Lincom Europa.

---. In preparation. ---. Vol. 3: Nominalklassifikation. Munich: Lincom Europa.

Smeets, Rieks. 1984. *Studies in West Circassian Phonology and Morphology.* Leiden: Hakuchi press.

----, ed. 1994. *The Indigenous Languages of the Caucasus, vol. 4: Northeast Caucasian Languages, Part 2.* Delmar, NY: Caravan.

Tsez beginnings

MARIA POLINSKY
University of California, San Diego

1. Introduction[1]
1.1. Basic facts.

Tsez (Nakh-Daghestanian) is an ergative language in which the verb agrees in grammatical class with the absolutive argument, either the subject of an intransitive or object of a transitive clause.[2] This is illustrated in (1) and (2). In (1), the intransitive verb *-ik'a* 'go' agrees in class with the absolutive subject. In (2), the transitive verb *-išra* 'feed' takes agreement prefixes which are determined by the class of the absolutive object.[3]

(1) a. kid   y-ik'i-s
         girl.ABS.II II-go-PSTEV
         'The girl went.'
     b. ziya   b-ik'i-s
         cow.ABS.IIIIII-go-PSTEV
         'The cow went.'
(2) a. kid-bā   ziya   b-išer-si
         girl-ERG   cow.ABS.III   III-feed-PSTEV
         'The girl fed the cow.'
     b. ziy-ā   kid   y-išer-si
         cow-ERG   girl.ABS.II   II-feed-PSTEV
         'The cow fed the girl.'

A verb with a sentential complement in the absolutive position will show class IV agreement. This is illustrated by (3a) where the head of the sentential complement is a nominalized verb, and (3b) where the sentential complement is an infinitival clause.

(3) a. muˤalim-ā   kid-beq   [dunyal   gelmaču
         teacher-ERG   girl-SUPERESS   [world   round
         yoł-xosi-łi]   r-iyr-er-si
         be-PRS.PRT-NMLZ].IV IV-learn-CAUS-PSTEV
         'The teacher taught the girl that the earth is round.'
     b. [t'ekmabi   t'et'r-a]   r-igu   yoł
         [books.ABS.PL   read-INF].IV   IV-good   is
         'To read books is good.'

1.2. *Begin* and *Continue*: Unusual agreement and case-marking.

Two aspectual verbs that take infinitival complements, -*oqa* 'become; begin' and -*iča* 'continue', display a different pattern of agreement. Contrary to the expected (4a) and (5a), where they should agree with the infinitival complement in class IV, -*oqa* and -*iča* show agreement with the subject of the infinitival clause. If the embedded infinitive is transitive, this NP can appear either in the absolutive (4b), (5b) or in the ergative (4c), (5c).

(4) a. *kid   [ziya   b-išr-a]   r-oq-si
      girl.ABS.II [cow.ABS.III   III-feed-INF].IV   IV-begin-PSTEV
      'The girl began to feed the cow.'

   b. kid   ziya   b-išr-a   y-oq-si
      girl.ABS.II cow.ABS.III   III-feed-INF   II-begin-PSTEV
      'The girl began to feed the cow.'

   c. kid-bā   ziya   b-išr-a   y-oq-si
      girl-ERG   cow.ABS.III   III-feed-INF   II-begin-PSTEV
      'The girl began to feed the cow.'

(5) a. *kid   ziya   b-išr-a   r-iči-s
      girl.ABS.II [cow.ABS.III   III-feed-INF].IV   IV-continue-PSTEV
      'The girl continued to feed the cow.'

   b. kid   ziya   b-išr-a   y-iči-s
      girl.ABS.II cow.ABS.III   III-feed-INF   II-continue-PSTEV
      'The girl continued to feed the cow.'

   c. kid-bā   ziya   b-išr-a   y-iči-s
      girl-ERG   cow.ABS.III   III-feed-INF   II-continue-PSTEV
      'The girl continued to feed the cow.'

The unusual constructions with -*oqa* and -*iča* raise the following question: what syntactic structures are associated with the different uses of these verbs? It is known that languages often have different aspectual verbs depending on transitivity, animacy and volitionality of the subject, and the ability of taking both nominal and sentential arguments (Perlmutter 1970; Newmeyer 1975; ter Meulen 1990, among others). In what follows, I will present evidence that Tsez also maintains the distinction between two types of *begin* and *continue*.[4] I will argue that in one of their uses *begin* and *continue* are Raising verbs; in the other use, the two verbs have a more unusual syntax, which I will call Backward Control. After discussing the two construction types associated with the two uses of *begin* and *continue* in Tsez, I will outline further questions raised by their syntax.

2. *Begin* and *Continue* as Raising Verbs

For sentences such as (4b) and (5b), I would like to claim that their structure involves subject-to-subject raising,[5] schematically shown in (6b) for transitive infinitives and (7b) for intransitives:[6]

(6)   a.   kid   ziya   b-išr-a   y-oq-si

                girl.ABS.II   cow.ABS.III   III-feed-INF   II-begin-PSTEV

                'The girl began to feed the cow.' (=(4b))

    b. kid_i [t_i ziya   b-išr-a]   y-oq-si

                girl.ABS.II   cow.ABS.III   III-feed-INF   II-begin-PSTEV

(7)   a.   kid   y-ik'-a   y-oq-si

                girl.ABS.II II-go-INF   II-begin-PSTEV

                'The girl began to leave.'

    b. kid_i [t_i y-ik'-a]   y-oq-si

                girl.ABS.II   II-go-INF   II-begin-PSTEV

Several arguments prove that the Raising analysis is correct.

2.1. Agreement

The Raising analysis immediately accounts for the agreement facts. Recall that Tsez verbs must agree with the absolutive argument, either intransitive subject, or direct object of transitive. Under Raising, the matrix verb agrees with the absolutive NP which raises to subject from the infinitival clause, and the embedded infinitive agrees with the absolutive NP in the infinitival clause. This is particularly clear in cases such as (4b) and (5b), where the embedded verb is transitive, and the two absolutive arguments belong to distinct noun classes.

2.2. Case-marking

The Raising analysis of *begin* and *continue* also explains why the subject of the embedded transitive clause appears in the absolutive and not the ergative case. In Tsez, all intransitive verbs take the absolutive subject (in other words, there is no split ergativity). As subject raising verbs, *begin* and *continue* are intransitive, the raised subject of these verbs must appear in the absolutive.

2.3. Relativization

As a result of Raising, the absolutive NP becomes part of the clause where -*oqa* or -*iča* is the predicate. The raised NP must be able to participate in the syntax of that clause and interact with other constituents of this clause. Passivization and relativization are among standard arguments that the raised NP bears a grammatical function in the target clause. Tsez does not have passives. As for relativization, it confirms that the absolutive NP undergoes Raising. Compare (6c), repeated below, and (8b):

(8) a. kid$_i$ [t$_i$ ziya b-išr-a] y-oq-si

   girl.ABS.II cow.ABS.III III-feed-INF II-begin-PSTEV

   'The girl began to feed the cow.' (=(6c))

  b. [$_{Rel}$ t$_i$ [ziya   b-išr-a] y-āq-ru] kid$_i$

     cow.ABS.III III-feed-INF II-begin-PST.PRT girl

   'the girl that began to feed the cow.'

2.4. Scrambling

 The raised NP can scramble freely with the constituents of the higher clause, as shown by (9). To follow the examples in (9) and (10), it is crucial to know that although Tsez is head-final, it permits postverbal NPs in the finite declarative clause. Thus, (9c) is well-formed.

(9) a. ħuł   kid$_i$ [t$_i$ ziya   b-išr-a]   y-oq-si

   yesterday girl.ABS.II cow.ABS.III III-feed-INF II-begin-PSTEV

  b. kid$_i$ ħuł   [t$_i$ ziya b-išr-a] y-oq-si

   girl.ABS.II yesterday cow.ABS.III III-feed-INF II-begin-PSTEV

  c. ħuł   [t$_i$ ziya   b-išr-a] y-oq-si kid$_i$

   yesterday cow.ABS.III III-feed-INF II-begin-PSTEV girl.ABS.II

   'Yesterday, the girl began to feed the cow.'

 In the meantime, the constituents of the infinitival clause cannot scramble with the constituents of the higher clause, as shown by the ungrammatical examples in (10). Unlike finite clauses, infinitival clauses must be strictly head-final—compare the well-formed (9c) with the ungrammatical (10c), where an NP follows the infinitive:

(10) a. *ħuł   ziya   kid   b-išr-a   y-oq-si

   yesterday cow.ABS.III girl.ABS.II III-feed-INF II-begin-PSTEV

  b. *ħuł   kid b-išr-a   y-oq-si   ziya

   yesterday girl.ABS.II III-feed-INF II-begin-PSTEV cow.ABS.III

  c. *ħuł   kid [b-išr-a ziya]   y-oq-si

   yesterday girl.ABS.II III-feed-INF cow.ABS.III II-begin-PSTEV

   'Yesterday the girl began to feed the cow.'

2.5. Clitic placement [7]

 Tsez has certain second position clitics which can appear only in the matrix clause. One such clitic is –*uy* that agrees in grammatical class with the absolutive. The clitic –*uy* roughly means 'indeed, definitely', which suggests that it may function as a validator indicating the evidential status of the reported event. This clitic can follow the raised absolutive NP, which confirms that this NP is in the

matrix clause. This is illustrated in (11a); (11b) shows that *–uy* cannot appear in the embedded clause, even if the second-position criterion is met.[8]

(11) a. kid   y-uy   [t$_i$ ziya   b-išr-a]   y-oq-si
      girl.ABS.II II-VAL   cow.ABS.III   III-feed-INF   II-begin-PSTEV
      'The girl indeed began to feed the cow.'

  b. *[ziya   b-uy   b-išr-a]   y-oq-si   kid
      cow.ABS.III   III-VAL   III-feed-INF   II-begin-PSTEV   girl.ABS.II
      'The girl began indeed to feed the cow.'

2.6. Absence of selectional restrictions

It is well known that raising verbs do not impose any semantic constraints on the raised NP—this accounts for the raising of parts of idioms or inanimate nouns. In Tsez, parts of idiomatic expressions occur with *begin* and *continue*, which is another indication that these are raising verbs. Compare (12a) and (12b), where the idiom is 'darkness eats the sun':

(12) a. t'ont'oħ-ā   buq   b-ac'-xo
      darkness-ERG sun.ABS.III   III-eat-PRS
      'The sun has eclipsed.'

  b. t'ont'oħu$_i$ [t$_i$ buq   b-ac'-a]   b-āq
      darkness.ABS.III   sun.ABS.III   III-eat-INF III-begin-FUT
      'The sun will begin to eclipse.'

In (13), the idiomatic expression involves the absolutive NP, and evidence for Raising in (13b) comes from the scrambling facts—the raised NP and the infinitival clause are on different sides of the verb *-oqa*.

(13) a. eniw-s   debe-ʎ'   rok'u   r-ay-x
      mother-GEN   2SG-SUPERESS heart.ABS.IV   IV-come-PRS
      'The mother relies on you (lit.: The mother's heart comes upon you.).'

  b. [t$_i$ debe-ʎ'   r-ay-a]   r-oq-si   eniw-s
      2SG-SUPERESS IV-come-INF   IV-begin-PSTEV   mother-GEN
      rok'u$_i$
      heart.ABS.IV
      'The mother began to rely on you.'

2.7. Cyclicity

Subject-to-subject raising can be cyclic, that is, the raised subject may undergo raising into an even higher clause (super-raising). Such cyclicity is observed for Tsez as well. In (14), the NP 'girl' first raises from the subject of the

embedded infinitival clause to the subject of *continue* and then into the next clause, to the subject of *begin*.[9]

(14) kid$_i$ [t$_i^2$ [t$_i^1$ xabar   es-a]   y-ic-a]   y-oq-si
       girl.ABS.II   story.ABS tell-INF   II-continue-INF   II-begin-PSTEV
       'The girl began to continue telling the story.'

If the Raising analysis did not apply, it would be impossible to account for the agreement on both continue and begin in (14). Under the Raising analysis, the agreement on the finite verb is trivially determined by the super-raised NP, and the agreement on the intermediary verb indicates that the NP *kid* has a representation in that clause too.

To summarize, the following arguments indicate that *begin* and *continue* can be raising verbs. First, the raised NP functions grammatically as a constituent of the clause into which it raises. This is shown by agreement facts, case marking, relativization, scrambling, and the interaction with the validator clitic -*uy*. Second, as raising verbs, *begin* and *continue* do not impose any selectional restrictions on the raised NP, which can be an idiom chunk, an inanimate noun (see also below), and a weakly quantified noun (example (27) below). Finally, one raising verb may be embedded under the other, leading to super-raising.

The use of *begin* and *continue* as raising verbs is cross-linguistically well-attested—see Perlmutter (1970) for English, McCloskey (1983) for Irish, and Kibrik (1987, 1999:492-514) for other Nakh-Daghestanian languages. Unlike some languages but rather like English, Tsez does not permit *begin* and *continue* to occur in the construction without Raising—the structures in (15a, b) are ungrammatical:

(15) a. *expletive* [kid-bā   ziya   b-išr-a]   r-oq-si
              girl-ERG   cow.ABS.III   III-feed-INF   IV-begin-PSTEV
          'The girl began to feed the cow.'
      b. *expletive*   [kid   y-ik'-a]   r-oq-si
              girl.ABS.II II-go-INF   IV-begin-PSTEV
          'The girl began to leave.'

I believe that the ungrammaticality of (15a, b) can be explained by the restricted use of expletives in Tsez, but a full discussion of this question is beyond the scope of this paper.

3. The Other Use of *Begin* and *Continue*

Let us now turn to the other structure with *begin* and *continue*, repeated here for convenience:

(4) c. kid-bā ziya   b-išr-a y-oq-si
    girl-ERG cow.ABS.IIIIII-feed-INF II-begin-PSTEV
    'The girl began to feed the cow.'

Two facts are intriguing about this construction. First, the aspectual verb agrees in grammatical class with a semantic argument of the embedded infinitival clause. Second, it seems to agree with the ergative NP, an otherwise unattested possibility in Tsez. Even if the ergative NP *kidbā* raised into the matrix clause, it's ability to trigger class agreement on *–oqa* would still be mysterious. However, the raising analysis is completely implausible—constituency tests show that the ergative NP remains in the embedded infinitival clause.

3.1. Overt subject in the embedded clause: Evidence from constituency
The ergative NP can scramble with the constituents of the embedded infinitival clause, as shown by (16a). This indicates that it is a constituent of the embedded infinitival clause. Now recall that finite clauses are not strictly verb-final. As the ergative NP and the remainder of the infinitival clause form a single constituent, this constituent can either precede or follow the finite aspectual verb. This is demonstrated by (16b) and (16c):

(16) a. ħuł   [ziya   kid-bā b-išr-a] y-oq-si
     yesterday cow.ABS.III girl-ERG III-feed-INF II-begin-PSTEV
     'Yesterday the girl began to feed the cow.'
   b. ħuł   [kid-bā ziya   b-išr-a] y-oq-si
     yesterday girl-ERG cow.ABS.III III-feed-INF II-begin-PSTEV
     'Yesterday the girl began to feed the cow.'
   c. ħuł   y-oq-si [kid-bā ziya   b-išr-a]
     yesterday II-begin-PSTEV girl-ERG cow.ABS.III III-feed-INF
     'Yesterday the girl began to feed the cow.'

More crucially, the ergative NP cannot scramble with the constituents of the higher clause, containing the aspectual verb. This is illustrated by the ungrammatical (17a, b):

(17) a. *ħuł   kid-bā$_i$ y-oq-si   [t$_i$ ziya
     yesterday girl-ERG II-begin-PSTEV cow.ABS.III
    b-išr-a]
    III-feed-INF
   b. *kid-bā$_i$ ħuł [t$_i$ ziya   b-išr-a] y-oq-si
     girl-ERG yesterday cow.ABS.III III-feed-INF II-begin-PSTEV
     'Yesterday the girl began to feed the cow.'

3.2. Silent element in the matrix clause

In addition to the indications that the ergative NP is inside the infinitival clause, there are several arguments that the higher clause contains a silent absolutive NP co-indexed with the lower subject.

3.2.1. Depictives

Tsez depictives such as *sisxoli* 'alone', *xizāz* 'last', *adāz* 'first' have two crucial properties: 1) they can be oriented towards the subject regardless of its case-marking or towards the absolutive; 2) they can follow the word they modify or they can be dislocated to the right—but not to the left—of that word. For example, in (18a), *sisxoli* is oriented towards 'girl' but not 'cow'; in (18b), its orientation is ambiguous. In the meantime, (18c) is ungrammatical because the depictive precedes the word towards which it is oriented.

(18) a. kid-bā   sisxoli ziya   b-išer-si
        girl-ERG   alone   cow.ABS   III-feed-PSTEV
        'The girl$_i$ fed the cow$_j$ alone$_{i/*j}$.'

    b. kid-bā   ziya   sisxoli b-išer-si
        girl-ERG   cow.ABS   alone   III-feed-PSTEV
        'The girl$_i$ fed the cow$_j$ alone$_{j>i}$.'

    c. *sisxoli   kid-bā   ziya   b-išer-si
        alone   girl-ERG   cow.ABS   III-feed-PSTEV

Now compare the following pair of sentences:

(19) a. *sisxoli   kid$_i$ [t$_i$ ziya   bišr-a]   y-oq-si
         alone   girl.ABS   cow.ABS   feed-INF   II-begin-PSTEV
         'The girl$_i$ began to feed the cow alone$_i$.'

    b. sisxoli   kid-bā   ziya   bišr-a   y-oq-si
        alone   girl-ERG   cow.ABS   feed-INF   II-begin-PSTEV
        'The girl$_i$ began to feed the cow alone$_i$.'

    c. *ec$_i$* sisxoli   [kid-bā$_i$ ziya   bišr-a]   y-oq-si
            alone   girl-ERG   cow.ABS   feed-INF   II-begin-PSTEV

Example (19a), where the depictive precedes the raised NP 'girl' is ungrammatical—the depictive must appear on the right of the noun it modifies. The only way to explain why (19b) is grammatical is to assume that there is a silent element preceding *sisxoli* and co-indexed with the ergative NP in the infinitival clause. This means that (19b) should be represented as (19c). The silent element in the matrix clause is in the absolutive position; it matches the embedded ergative in grammatical class and determines the agreement on the verb –*oqa*.

3.2.2. Long-distance agreement

The second argument for a silent element in the clause with –*oqa* or –*iča* comes from the so-called LONG-DISTANCE AGREEMENT. Under long-distance agreement in Tsez, a verb may agree with a constituent inside its clausal complement (Polinsky 1999; Potsdam and Polinsky 1999; Polinsky and Potsdam 1999a). A crucial fact about the Tsez long-distance agreement is that it cannot cross multiple clause boundaries—the higher verb can only agree with the absolutive in the next embedding but not lower. For example, in (20a), the higher verb 'know' agrees in class with the embedded absolutive 'cow', but in (20b), where there is a third level of embedding, the intermediary verb 'know' agrees with the embedded absolutive, but the higher verb 'like' cannot agree with it.

(20) a. eni-r   [kid-bā   ziya
          mother-DAT   girl-ERG   cow.ABS.III
          b-išer-xosi-ɫi]   b-iy-x
          III-feed-PRS.PRT-NMLZ   III-know-PRS
          'The mother knows that the girl is feeding the cow.'
      b. obi-r   [[kid-bā   ziya   b-išer-xosi-ɫi]
          father-DAT   girl-ERG   cow.ABS.III   III-feed-PRS.PRT-NMLZ
          eni-r   b-iy-xosi-ɫi]   r-eti-x/*b-eti-x
          mother-DAT   III-know-PRS.PRT-NMLZ   IV-like-PRS/*III-like-PRS
          'The father likes that the mother knows that the girl is feeding the cow.'

When a construction with *begin* or *continue* is embedded under a higher verb, this verb can show long-distance agreement matching the grammatical class of the overt ergative NP. For example, in (21), the verb 'know' can show class II agreement, thus agreeing with the NP 'girl'.

(21) eni-r   [ec_i [kid-bā$_i$ ziya   b-išr-a]
      mother-DAT   girl-ERG   cow.ABS.III   III-feed-INF
      y-āq-ru-ɫi]   y-iy-x
      II-begin-PST.PRT-NMLZ   II-know-PRS
      'The mother knows that the girl began to feed the cow.'

Given the evidence presented earlier, the ergative NP is embedded too deep for the verb 'know' to agree with it (let alone that the ergative-triggered agreement is otherwise impossible in Tsez). The long-distance agreement on 'know' can be accounted for if there is a silent element in the clause with the verb *begin*, because this clause is immediately dominated by the matrix clause in (21).

To summarize the arguments presented so far, I have shown that *begin* and *continue* form a construction in which the embedded subject (absolutive or ergative) remains inside the infinitival clause. This subject is co-indexed with a silent absolutive which is the argument of *begin/continue* and which triggers the

agreement on the aspectual verb.[10] Schematically, the proposed structure is illustrated in (22b); (22a) repeats the raising structure for *begin* and *continue*:

(22) a. NP_i $[t_i\ VP_{inf}]$ *begin/continue*
      b. ec_i $[Subject_i\ VP_{inf}]$ *begin/continue*

3.3. Status of the infinitival clause

Under Raising, *begin* and *continue* are used intransitively. What about their second use? The structure in (22b) has to be intransitive as well, for several reasons. First, the silent element co-indexed with the embedded subject is in the absolutive position, otherwise it would not be able to trigger agreement. If so, it is either the intransitive subject or the direct object; the latter would mean that the infinitival complement should be identified with the transitive subject, a highly implausible situation. This leaves us to conclude that the silent element is the absolutive intransitive subject. Next, the verbs *begin* and *continue* can form imperatives, which indicates that they take a second person animate subject (see Perlmutter 1970: 113 for a similar argument). Imposing selectional restrictions on the subject is incompatible with the syntax of Raising, which means that the imperative of *begin/continue* has to be formed from (22b), not (22a). For certain verb stems, including *–oq-*, Tsez has a morphological contrast between imperatives of intransitives (formed with the zero suffix) and imperatives of transitives, which take the suffix *–o*. If the aspectual verbs in (22b) were transitive, we would expect the imperative of *–oqa* to be *–oqo*, if intransitive— *–oq*. Only the latter form is actually attested, showing again that *begin* is intransitive:

(23)   ziya   bišr-a   y-oq/*y-oq-o
        cow   feed-INF   II-begin.IMPER.INTR/ II-begin.IMPER.TR
        'Begin to feed the cow.' (addressing a woman)

This evidence suggests that *begin* and *continue* are intransitive in their second use. The infinitival complement in (22b) should then be interpreted as an adjunct complement clause. Two arguments support the proposed adjunct status of this complement. First, the infinitival complement can be omitted, which is typical of adjuncts. Second, the aspectual verbs *begin* and *continue* cannot be used with non-clausal objects. Different verbs have to be used to express something like "They began the project" (*baybik boda* 'begin (lit.: beginning make)') or "They continued the project" (*-exorik'a* 'prolong, continue').

Having demonstrated that the second use of *begin* and *continue* involves a structure as in (22b), I would like to propose that in this structure the verbs act as subject control verbs.

4. Backward Control

4.1. Subject control: Identity between the silent element and the lower subject

The higher absolute subject and the subject in the infinitival clause cannot be expressed simultaneously. This is shown by (24):

(24)   *kid$_i$ [ziya   kid-bā$_i$ bišr-a]   y-oq-si
        girl.ABS   cow.ABS   girl-ERG   feed-INF   II-begin-PSTEV
        'The girl began to feed the cow.'

It is equally impossible for *begin* and *continue* to occur in structures where their subject is different from the subject of the embedded infinitival clause—this is demonstrated by the ungrammatical (25a, b):

(25)   a.   *kid   [eniy   y-ik'-a]   y-oq-si
          girl.ABS   mother.ABS   II-go-INF   II-begin-PSTEV
          'The girl started the mother's departure.'
    b.   *kid   [eniy-ā   ziya   bišr-a]   y-oq-si
          girl.ABS   mother-ERG   cow.ABS   feed-INF   II-begin-PSTEV
          'The girl started the mother's feeding the cow.'

The requirement that the two subjects be identical further suggests that *begin* and *continue* are subject control verbs. Direct evidence for that comes from selectional restrictions.

4.2. Subject control: Selectional restrictions

Control verbs are known to impose selectional restrictions on their arguments. Inanimate NPs, idiom chunks, and weakly quantified (existential) NPs require the Raising construction. The animacy restriction was already mentioned above, in the discussion of the imperatives. Now compare the well-formed Raising sentence in (26a) with the ungrammatical (26b). The contrast between (26a) and (26b) illustrates both the animacy restriction and the fact that idiom chunks can undergo Raising but are impossible in the other construction with *begin* and *continue*.

(26)   a.   t'ont'oħu$_i$ [t$_i$ buq   b-ac'-a]   b-āq
          darkness.ABS.III   sun.ABS.III   III-eat-INF   III-begin.FUT
          'The sun will begin to eclipse.' (=(12b))
    b.   *ec$_i$ [t'ont'oħ-ā$_i$ buq   b-ac'-a]   b-āq
          darkness-ERG   sun.ABS.III   III-eat-INF   III-begin.FUT
          'The sun will begin to eclipse.'

Next, existential NPs cannot occur in the structure in (22b). This can be demonstrated using the language-internal contrast between restrictive and non-restrictive adjectives. An NP modified by a non-restrictive adjective must have a

weak (existential) construal. Nouns modified by non-restrictive adjectives must occur in the Raising construction (27a) and cannot occur in the control construction with *begin* or *continue* (27b):

(27) a. aluk'a-ø   aɣi-bi$_i$ [t$_i$ k'eneč'yabi   et'-a]   r-oq-xo
          white-NR bird-ABS.PL egg.ABS.PL   pour-INF PL-begin-PRS
          'White birds begin to lay eggs.'
      b. *ec$_i$ [aluk'a-ø aɣi-za$_i$ k'eneč'yabi   et'-a]   r-oq-xo
          white-NR bird-ERG.PL   egg.ABS.PL   pour-INF PL-begin-PRS
          'White birds begin to lay eggs.'

Thus, the selectional restrictions indicate that the verbs *begin* and *continue* function as subject control verbs.[11] They meet, in a peculiar way, the well-known characteristics of obligatory control (there must be a controller; the controller must be the subject; a lexical NP cannot be substituted for the silent element). However, the striking fact about *begin* and *continue* as control verbs is that they delete the matrix (higher) subject under identity with an embedded subject. I will refer to this property of *begin* and *continue* as Backward Control (Farrell 1995).

4.3. On Backward Control

Backward Control subsumes the following characteristics: 1) both the aspectual verb (*begin* or *continue*) and the infinitive take the referent of the subject as their argument; 2) the aspectual verb assigns its subject the semantic role agent; 3) the subject of the aspectual verb and the subject of the infinitive are identical; 4) the subject of the aspectual verb cannot be expressed by a lexical NP and must be expressed by a null pronominal; 5) the aspectual verb is intransitive—it takes a subject and an adjunct infinitival clause.

These properties account for a number of features observed in the Tsez control structures with *begin* and *continue*. In particular, properties 1 and 2 explain the selectional restrictions observed in the control structures. Property 2 explains the formation of imperatives and may also be linked to various semantic effects associated with the control structure (see fn. 11). Property 3 accounts for the ill-formed sentences where the two subjects do not match. Property 3 also shows that case identity of the higher NP and lower NP does not matter as long as they are both subjects. Property 4 implies that the subject of the aspectual verb must be expressed by *pro*. Property 5 explains why the agent of *begin* or *continue* is in the absolute case and thus triggers the agreement in the aspectual clause. Property 5 also explains why the infinitival complement seems to lack object properties.

There are several mentions of Backward control (or as it was called earlier, Counter-Equi) in the literature. Backward Object Control has been proposed to account for the syntax of causatives in Japanese (Harada 1973; Kuroda 1978) and in Brazilian Portuguese (Farrell 1995). A Backward Control analysis has also been proposed for the adjunct clauses headed by *tokoro* in Japanese (Harada

1973), and possibly for infinitival control structures in Jacaltec (Craig 1977: 257-8, 323-5). Descriptive studies suggest that at least two other Nakh-Daghestanian languages may have Backward Control: Tsaxur seems to use it with the *verbs begin, stop, hurry*, and *be afraid* (Kibrik 1999: 499-500), and Hunzib, a language closely related to Tsez, seems to apply Backward Control with the verb 'can' (van den Berg 1995: 129-30).

In all these cases, just as in Tsez, Backward Control is an option for particular verbs. The relevant verbs include aspectual, volitional, and causative predicates which form a subset of the aspectual, volitional, and causative predicates within a given language. For instance, in Tsez, aspectual predicates other than *begin* and *continue* either form clause union structures (Polinsky and Potsdam 1999b) or embed clausal complements without argument deletion. In addition to Backward Control, Tsez also has regular control structures with such verbs as *promise* and *decide*.

The goals of this paper do not include a comprehensive account of Backward Control. I have demonstrated that it is empirically cogent and would like to conclude with the main issues that this phenomenon raises for linguistic theory. First, Backward Control is problematic for existing theories of control (Williams 1980; Manzini 1983. Sag and Pollard 1991) because it fails to meet the binding criteria on control (see also Farrell 1995: 124). It is not entirely clear if the movement theory of control (Hornstein 1999) can handle Backward Control either.

Whatever theory of control is adopted, it must motivate the choice between regular (subject and object) control and Backward Control. The data from Tsez, Brazilian Portuguese, and Japanese *tokoro*-clauses suggest that the two control types are mutually exclusive and their choice is determined by a specific lexical item or a specific construction type.[12] Some other Japanese data suggest that the two types of control may partially overlap (Kuroda 1978).

Backward Control seems a more restricted option than regular control, and its existence also raises a typological question: What structural properties of a language determine that it will or may have Backward Control? At this point, there is no sufficient theory or empirical base to answer this question.

5. Conclusion

In this paper, I have examined two constructions with the Tsez aspectual verbs *begin* and *continue* which take infinitival complements. I have shown that one of these constructions involves Subject-to Subject Raising out of the infinitival clause. The other construction involves a more unusual phenomenon of Backward Control. Under Backward Control, the subject of the embedded clause remains in the embedded clause and must be expressed, and the subject of the aspectual verb must be deleted under coreference with the embedded subject of the infinitive. I have proposed that, when they appear in a Backward Control structure, *begin* and *continue* are still intransitive, taking an agentive subject and an adjunct infinitival complement.

The difference in two structures clearly has semantic correlates (the Raising construction does not impose any restrictions on its subject; and can also have a motion verb interpretation). As I have shown above, inanimate subjects can occur only in the Raising construction. For those NPs that can occur in either construction, further work is needed to determine whether or not semantic factors can predict the choice of a particular structure.

Notes

The following abbreviations are used: ABS—absolutive, CAUS—causative, DAT—dative, ERG—ergative, FUT—future, GEN—genitive, IMPER—imperative, INF—infinitive, INTR—intransitive, NMLZ—nominalizer, NON.RESTR—non-restrictive, PL—plural, PRS—present, PRS.PRT—present participle, PST.PRT—past participle, PSTEV—past evidential, SUPERESS—superessive, TR—transitive, VAL—validator. The numerals I through IV indicate noun classes.

[2] For the details of Tsez agreement, see Polinsky and Comrie 1999. To follow the discussion below, the reader should bear in mind that agreement prefixes do not attach to consonant-initial verbs and a small subset of vowel-initial verbs. This restriction is due to the phonotactic rules of Tsez, which I don't review here.

[3] Reflexivization, coreferential deletion across clauses and causative formation prove that the ergative NP is the subject of a transitive clause, something I assume as a given for the purposes of this paper.

[4] I am assuming that Tsez has a single verb *begin* and a single verb *continue*, each of which has two distinct uses. Nothing hinges on this assumption.

[5] Rosenbaum 1967; Postal 1974; Ruwet 1991; Langacker 1995, and many others.

[6] Here and below, I will be using only the verb *-oqa* for illustration; the behavior of *–iča* is identical.

[7] A similar phenomenon is found in Quechua (Cole and Hermon 1979: 79-80).

[8] The contrast between (11a) and (11b) also shows that the structure is biclausal.

[9] Aside from *begin* and *continue*, there are no other raising verbs in Tsez. These verbs do not exactly match semantically, which makes possible examples of Super-raising rather odd.

[10] Until now, I have been using transitive infinitival clauses because those manifest an overt contrast between the intriguing structure in (22b) and the Raising structure in (22a)—only the former features an ergative NP. However, the two structures also make different predictions for embedded intransitives. In (22a), the semantic argument of the embedded clause is the grammatical subject of the higher clause, thus, separate from the infinitival clause. In (22b), the respective absolutive argument is inside the infinitival clause. Recall that finite clauses in Tsez are not strictly head-final. This predicts that the absolutive NP and the infinitival clause can appear on different sides of *begin* or *continue* in the Raising structure (22a), but not in (22b). This prediction is borne out:

(i) a. ok'o$_i$ [t$_i$ b-oš-a]   b-oq-si
         nail.ABS.III   III-bend-INF   III-begin-PSTEV

     b. [t$_i$ b-oš-a]   b-oq-si   ok'o$_i$
               III-bend-INF   III-begin-PSTEV   nail.ABS.III

     c. b-oq-si   [t$_i$ b-o_-a]   ok'o$_i$
         III-begin-PSTEV   III-bend-INF   nail.ABS.III
                       'The nail began to bend.'

The control structure should not allow orders where the absolutive NP is separated from the infinitival clause or where it follows the infinitive—recall that infinitival complements must be verb-final. This rules out orders such as in (i-b) and (i-c).

[11] The syntax of obligatory control also has semantic correlates. First, starting with the fact that the silent NP in the control structure has the semantic role Agent, certain predictions can be made with regard to the

agentivity and volitionality of the subject and the choice between Raising and Control. If the subject of *begin/continue* is clearly volitional, the control structure has to be used (e.g., with the agentive adverbials 'on purpose', 'knowingly', etc.). Second, *-oqa* receives slightly different semantic interpretations in the Raising and in the Control structures—only under Raising does one find the meaning of motion superimposed on the meaning 'begin'. For instance, (4b) is translated as 'the girl set out/went out to feed the cow'. At this point, I have no explanation for this difference in interpretation.

[12] Farrell (1995: 125) motivates the choice in terms of event structure.

References

Cole, Peter, and Gabriella Hermon. (1979). Subject to Object Raising in an EST framework: Evidence from Quechua. *Studies in the Linguistic Sciences* **9**, 65-90.

Craig, Colette G. (1977). *The Structure of Jacaltec*. Austin: University of Texas.

Farrell, Patrick. (1995). Backward Control in Brazilian Portuguese. *ESCOL* **11**, 116-27.

Harada, S. I. (1973). Counter Equi NP Deletion. *University of Tokyo. Research Institute of Logopedics and Phoniatrics. Annual Bulletin* **7**, 113-47.

Hornstein, Norbert. (1999). Movement and Control. *Linguistic Inquiry* **30**, 69-96.

Kibrik, Alexandr E. (1987). Constructions with clause actants in Daghestanian languages. *Lingua* **71**, 133-178.

Kibrik, Alexandr E., ed. (1999). *Elementy caxurskogo jazyka v tipologičeskom osveščenii*. Moscow: MGU.

Kuroda, S.-Y. (1978). Case-marking, canonical sentence patterns, and Counter-Equi in Japanese. In J. Hinds, ed. *Problems in Japanese Syntax and Semantics*. Tokyo: Kaitakusha, 30-51.

Langacker, Ronald W. (1995). Raising and transparency. *Language* **71**, 1-62.

Manzini, Rita. (1983). On Control and Control Theory. *Linguistic Inquiry* **14**, 421-46.

McCloskey, James. (1983). Raising, subcategorization and selection in Modern Irish. *Natural Language and Linguistic Theory* **1**, 441-85.

Newmeyer, Frederick J. (1975). *English Aspectual Verbs*. The Hague: Mouton.

Perlmutter, David M. (1970). The two verbs *Begin*. In R. Jacobs and P. Rosenbaum, eds. *Readings in English Transformational Grammar*. Waltham, Mass.: Blaisdell, 107-19.

Polinsky, Maria. (1999). Variation in complementation constructions: Long-distance agreement in Tsez". In K. Horie and M. Verspoor, eds. *Studies in Complementation*. Amsterdam: John Benjamins.

Polinsky, Maria, and Bernard Comrie. (1999). Agreement in Tsez. *Folia Linguistica* **23**, 1-52.

Polinsky, Maria, and Eric Potsdam. (1999a). *Long-Distance Agreement.* Ms, UCSD and Yale University.

Polinsky, Maria, and Eric Potsdam. (1999b). Cross-linguistic view of long-distance agreement. Paper presented at the Third International Conference of Association of Linguistic Typology. Amsterdam, August 1999.

Postal, Paul M. (1974). *On Raising: One Rule of English Grammar and Its Theoretical Implications.* Cambridge, Mass., MIT Press.

Potsdam, Eric, and Maria Polinsky. (1999). Long-distance agreement in Tsez. *WCCFL* 18.

Rosenbaum, Peter. (1967). *The Grammar of English Predicate Complement Constructions.* Cambridge, Mass., MIT Press.

Ruwet, Nicolas. (1991). *Syntax and Human Experience.* Chicago: University of Chicago Press.

Sag, Ivan, and Carl Pollard. (1991). An integrated theory of complement control". *Language* **67**, 63-113.

ter Meulen, Alice. (1990). English aspectual verbs as generalized quantifiers. *NELS* **20**, 378-90.

Van den Berg, Helma. (1995). *A Grammar of Hunzib.* München: Lincom Europa.

Williams, Edwin. (1980). Predication. *Linguistic Inquiry* **11**, 203-38.

Kannada gerund in adnominal positions: a functional perspective

MIRJAM FRIED
University of California, Berkeley

1. Introduction

This paper examines as yet unresolved distributional patterns displayed by three non-finite verb forms in Kannada. One of the forms, shown in (1), is usually referred to as a gerund or verbal noun (Spencer 1914, Nadkarni 1972, Andronov 1979, Sridhar 1990). The form illustrated in (2) is a participle described variously as 'adjectival' (Shiffman 1983), 'adnominal' (Steever 1988), or 'relative' (Sridhar 1990). The form exemplified in (3) is also known under a variety of names: a participial noun (Spencer 1914, Andronov 1979), nominalized verb (Shiffman 1983), a verbal noun with pronominal head (Steever 1998), or a relative participle with pronominal head (Sridhar 1990).

(1) sekega:ladalli   *malaguv-ud-u*   sulab^{h}ava:gittu[1]
    hot:weather:LOC   sleep:NPST-G-NOM   easy:happen:PST:3SG:N[2]
    'Sleeping in hot weather was easy.'

(2) na:nu   aḍige   *ma:ḍuttiruv-a*   hengasannu   no:ḍidenu
    1SG:NOM   food   make:NPST-AP   woman:ACC   see:PST:1SG
    'I saw the woman that was cooking.'

(3) naanu   aḍige   *maaḍuttiruv-avan-annu*   nooḍidenu
    1SG:NOM   food   make:NPST-AN:SG:M-ACC   see:PST:1SG
    'I saw the one[male] who was cooking.'

The terminological variety alone suggests that there is no consensus about the status of these forms, structurally or functionally. Standard descriptions and grammars (e.g. Spencer 1914, Schiffman 1979, Sridhar 1990, Steever 1998) treat them as unrelated strategies for turning finite verbs into non-finite complements of various types. Steever 1998:149, for example, describes all three of them collectively as forming relative clauses, adverbial clauses, and nominal complements, without further differentiation that would pair individual forms with specific functions.

The purpose of this paper is to demonstrate that the three forms are functionally and morphosyntactically interrelated in ways that call for an explanation. For consistency, I will adopt the following terminology: GERUND for the form exemplified in (1), ANAPHORIC NOMINALIZATION for the form in (3), and, following Steever 1988, ADNOMINAL PARTICIPLE for the form in (2). I will argue that they constitute a network of distinct grammatical patterns with partial overlap in their function and morphosyntax. In order to establish the network, several factors will be considered: the morphological structure of the forms in (1) and (3), the syntactic and semantic function of all three forms, and the pragmatic and discourse constraints on (2) vs. (3).

2. Gerund

The point of departure will be a paradigmatic gap in the case marking of the gerund (G). Morphologically, G is formed from a tensed stem (past or non-past) by adding a nominalizing suffix followed by a case marker. Functionally, it is generally described as a verbal noun that behaves like any other noun in that it takes all the case inflections of regular nouns and instantiates nominal constituents in the full range of syntactic functions. In (1) above it is shown as a subject, in the examples below it instantiates a direct object (4a), an oblique in the dative (4b), and an oblique in the ablative (4c):

(4) a. avanu   maneya   oḷage   *ho:did-d-annu*   no:ḍide
        3SG:M:NOM   house:GEN   inside   go:PST-G-ACC   see:PST:1SG
        'I saw him going into the house.'
     b. kannaḍa   *kaliuv-ud-akke*   kaṣṭa
        Kannada:NOM   learn:NPST-G-DAT   difficult
        'Kannada is hard to learn.' (lit. 'Kannada is hard with respect to learning')
     c. i: hunnugaḷu   bandiruvudu   *o:ḍuv-ud-arind-*alla
        this blister:PL:NOM come:FPF:3SG:N run:NPST-G-ABL-NEG
        'These blisters aren't from running.'

As these examples demonstrate, G is used as a nominalized clause, as in (1) and (4a), or as a simple *nomen actionis* without complements, as in (4b) or (4c).

It is not clear, however, that G ever appears in the genitive. On the one hand, no Kannada verb governs a genitive and, therefore, genitive-marked nominals – gerundial or otherwise – cannot be found in argument positions. But G seems conspicuously absent even in adnominal environments that normally govern the genitive. Of particular interest is the construction illustrated in (5a-b) in which a noun-like entity (*me:le* 'on top of', *vis^haya* 'knowledge/about') requires a genitive-marked complement.[3] As (5c) shows, G is unacceptable in this environment and instead, the adnominal participle (AP) is used:

(5) a. me:jin-a   me:le
        table-GEN   on:top:of
        'on the table'
     b. na:nu   be:sa:yad-a   vis^haya   kalitenu
        1SG:NOM   farming-GEN   about   learn:PST:1SG
        'I learned about farming.'
     c. *bittuhooguv-a/*bittuhooguv-ud-ara* varege   baa
        leave:NPST-AP/*-G-GEN   to:point:of come:IMP
        'Come to the place of departure.'

The case marking on G thus appears to form the following paradigm (morphologically, the case markers are identical to those found on neuter sg. pronouns – a point to be addressed in section 5):

(6)   [Tensed stem – Nominalizer – Case]
          malaguv   -   *ud*   -   *u*   nominative
                                    -   *annu*   accusative
                                    -   *akke*   dative
                                    -   **ara/ (a)* genitive / (AP)
                                    -   *aralli*   locative
                                    -   *arinda*   ablative

This patterning raises minimally two questions: why should the gap exist and why should AP be a fitting replacement of the genitive. Adding to the fuzziness, Sridhar (1990: 96) offers the example shown here in (7) with the comment that '*munce* 'prior to' can take either genitive or dative'. What he refers to as genitive is morphologically an instance of AP, and his description is thus evidently motivated by functional considerations.

(7) tinnuv-ud-akke/*tinnuv-a* munce
    eat:NPST-G-DAT/-AP   prior:to
    'before eating'

To complicate matters further, the genitive-marked G sometimes *does* appear, as shown in (8b), alongside AP in (8a). When it does, however, it forces a different interpretation. Notice that the AP variant in (8a) is not synonymous with the genitive G in (8b): the gerundial form *baruvudara* introduces a resumptive relationship between the matrix clause and the embedded nominalization.

(8) a. ra:ma illige *baruv-a*   vis^{h}aya nanage   ma:tana:ḍabe:ku
        Rama here   come:NPST-AP about   1SG:DAT   talk:MOD
        'I want to talk about Rama's coming here.'
     b. ra:ma illige *baruv-ud-ara* vis^{h}aya nanage   ma:tana:ḍabe:ku
        Rama here   come:NPST-G-GEN about   1SG:DAT   talk:MOD
        '[Rama's coming here]$_i$; I want to talk about it$_i$.'

The resumptive flavor is not present in the constructions in which G instantiates an argument of the matrix verb, such as the ones in (1) or (4). It is, however, a property regularly associated with the anaphoric nominalization (AN), as will be discussed shortly; the introductory sentence in (3) above, for example, would be more accurately translated as 'the one that was cooking – I saw him'. Thus, in order to understand the nature of the case marking gap in G, we must first sort out these overlaps. That, in turn, requires that we establish the defining properties of both AP and AN and then examine their relationship to G.

3. Adnominal participle

Morphologically, AP consists of a tensed stem and the relational suffix *–a*. One of its functions is to form relative clauses, as illustrated in (2) above and in (9) below. All these examples contain a noun modified by an AP-headed participial clause.

(9) a. ni:nu   *tinnuv-a*   seebu
     2SG:NOM   eat:NPST-AP   apple:NOM
     'the apple that you will eat'

     b. na:yiyannu   *kond-a*   manuṣyannu   na:nu   dve:ṣisutte:ne
     dog:ACC   kill:PST-AP   man:ACC   1SG:NOM   hate:PRES:1SG
     'I hate the man that killed the dog.'

AP is also found in a wide range of 'adverbial' clauses, so called because they express various adverbial relations such as place, time, manner, comparison, measure, etc. The adverbial clauses are introduced by members of the set of indeclinable quasi-nominals shown in (5) and (7); additional examples are in (10).

(10) a. na:nu   a:gale: *koṭṭ-a*   me:le   ja:sti haṇa   he:ḷidanu
     1SG:NOM   already   give:PST-AP   on:top:of   more   money   ask:PST:3SG:M
     'He asked for more money in addition to what I already gave him.'

     b. avanu   hinde   *baruv-a*   oḷage   na:nu   mudukana:girutte:ne
     3SG:M:NOM   back   come:NPST-AP   inside   1SG:NOM   old:become:PRES:1SG
     'By the time he comes back I'll be an old man.'

These two uses represent two equally productive and functionally distinct patterns – modification in (9) and complementation in (10). On the surface, this distinction is made less prominent by the fact that both patterns employ the same embedding strategy, namely the use of AP, apparently because the overriding shared feature is the adnominal position. Both clause types are headed by a nominal of sorts: a full noun in the former and a quasi-nominal in the latter. In fact, there are cases of complementation headed by a full noun as well, thus giving further support to analyzing the functionally 'adverbial' heads as nominals. In (11), an AP clause instantiates the complement of the noun *sakti* 'ability':

(11)   manuṣyan-a   tondareyannu   *suttibaruv-a*   sakti
     man-GEN   hardship:ACC   come:around:NPST-AP   ability:NOM
     'man's ability to overcome hardship'

To summarize, AP is an indeclinable form that appears in syntactic slots headed by nominals and that expresses a proposition-based property of that nominal. It is only the nature of the nominal itself that determines how the embedded AP will be analyzed and interpreted, whether as a relative clause or as a complement. Under this view, the adverbial reading is no more than a semantically defined subtype of noun-governed complements.

4. Anaphoric nominalization

Syntactically, AN resembles G in that it also instantiates nominal constituents. The example in (3) above shows AN as an accusative-marked object, the following sentence illustrates a dative-marked experiencer argument:

(12)   bʰu:kampavannu   *anubʰavisid-avar-ige*   e:nu   ma:ḍabe:kendu
     earthquake:ACC   experience:PST-AN:PL:H-DAT   what   do:MOD:COMP
     gottide
     know:PPF:3SG:N

'[As regards] those who have experienced an earthquake, they know what to do.'

However, Kannada grammarians do not always classify this form as a type of verbal noun, focusing instead on its functional similarity with AP constructions. Sridhar (1990:142), for example, describes AN as a relative participle with a pronominal head. AN is thus treated merely as a subtype of AP relativization, a subtype reserved for those relative clauses that are headed by a pronoun instead of a full noun. On a closer look, however, there are significant differences between AN and AP and I will show that whatever similarity there is between the two forms, it does not extend beyond the fact that they both may involve a nominal somewhere in the sentence. More specifically, the comparison must be carried out on several fronts, addressing discourse-related differences, pragmatic differences, restrictions on the pronominal 'head', and the morphological structure of AN.

First we note that the AP-based relativization is not interchangeable with AN, but that each serves a different communicative function. This difference is illustrated by the examples in (13-14), where the (a) sentences contain AP-based relative clauses and the (b) sentences contain ANs. As indicated by the English translations, the overall effect of AN is that of a resumptive expression centered on the constituent that is modified by the relative clause in the AP versions – *manuṣya* 'man' in (13b) and *se:bʰannu* 'apple' in (14b).

(13) a. na:yiyannu   *kond-a*   manuṣyannu   na:nu   dve:ṣisutte:ne
     dog:ACC   kill:PST-AP   man:ACC   1SG:NOM   hate:PRES:1SG
     'I hate the man that killed the dog.'

   b. manuṣya   na:yiyannu   *kond-avan-annu*   na:nu   dve:ṣisutte:ne
     man:NOM   dog:ACC   kill:PST-AN:SG:M-ACC   1SG:NOM   hate:PRES:1SG
     '[As regards] the man that killed the dog – I hate him.'

(14) a. avanu   me:jina   me:le   *idd-a*   se:bʰannu   tindanu
     3SG:M:NOM   table:GEN   on:top:of   be:PST-AP   apple   eat:PST:3SG:M
     'He ate the apple that was on the table.'

   b. se:bʰannu   me:jina   me:le   *iddad-d-annu*   tindanu
     apple:NOM   table:GEN   on:top:of   be:PST-AN-ACC   eat:PST:3SG:M
     '[As regards] the apple on the table - he ate it.'

This difference in discourse function is signaled by several properties that are idiosyncratically associated with AN and absent in the AP-based clauses.

First of all, the nominals in question (*manuṣya, se:bʰannu*) are always placed in the sentence-initial position – a position reserved for topics or foregrounded material in Kannada. Second, the foregrounded nominal must appear in the nominative (or unmarked) case, regardless of its syntactic function in the main clause. Notice that in both (13) and (14) this nominal instantiates the object of the verbs *dve:ṣisu-* 'hate' and *tinnu-* 'eat', and as such should be in the accusative (cf. the AP versions). Thirdly, the foregrounded nominal is referenced by the pronominal morpheme in the AN variant, which also carries the expected case marking. And finally, the embedded non-finite clause in the AN-based versions is

pronounced with a comma intonation, functionally equivalent to the question-like, resumptive intonation at the end of the topicalized portion of the corresponding English sentences.

The AN construction is very productive and in colloquial speech apparently favored over the AP relativization. Nonetheless, there are contexts in which only AP is permitted, while in others, seemingly comparable ones, either form could be used. The relevant environment is found in embedded structures headed by the quasi-nominals, where the foregrounded portion is the entire embedded clause. The following sentences illustrate contexts in which both AP and AN are possible (15) and contexts in which AN is rejected (16):

(15) a. avanu   illige *baruv-a/baruv-ud-ara* munce manege   ho:danu
　　　　 3SG:M:NOM here   come:NPST-AP/-AN-GEN   prior home:DAT go:PST:3SG:M
　　　　 'He went home before coming here.' (AP)
　　　　 '[As regards] his coming here, he went home before that.'   (AN)

　　 b. avanu   illige *baruv-a/baruv-ud-ara* munce   na:nu   a:gale:
　　　　 3SG:M:NOM here   come:NPST-AP/-AN-GEN   prior   1SG:NOM already
　　　　 adara   bagge   ma:tana:ḍi   a:giruttade
　　　　 3SG:N:GEN about   speak:PPL   happen:FPF:1SG
　　　　 'I'll have already spoken about it before he comes.' (AP)
　　　　 '[As for] his coming here – by that time I'll have already spoken about it.' (AN)

(16) a. maneya   oḷage *baruv-a/*baruv-ud-ara* munce avaḷu   meṭṭanu
　　　　 house:GEN inside come:NPST-AP/*-AN-GEN prior 3SG:F:NOM shoes:ACC
　　　　 tegedaḷu
　　　　 take.off:PST:3SG:F
　　　　 'She took her shoes off before entering the house.'
　　　　 *'[As for] her coming in the house, she took her shoes off before that.'

　　 b. kelasakke   *ho:guv-a/*ho:gov-ud-ara* munce bekkige   u:ṭakoḍuvanu
　　　　 work:DAT go:NPST-AP/*-AN-GEN prior cat:DAT   feed:NPST:3SG:M
　　　　 'He'll feed the cat before going to work.'
　　　　 *'[As regards] his going to work, he'll feed the cat before that.'

In (15) it does not seem to matter whether the embedded clause is foregrounded (AN) or not (AP); the choice will, presumably, depend on the larger context. The unavailability of AN in (16), however, suggests that AN is sensitive to a certain degree of conventionalization in pairing sequences of events. The two events reported in (15a) or those in (15b) are independent of each other and their mutual relationship can, therefore, be constructed either way. In contrast, the events in (16a) or those in (16b) represent conventionally fixed sequences of events and the effect of using AN would be one of disconnecting and rearranging them as if they had nothing to do with each other. Put differently, the chronologically subsequent event in such conventionalized pairs cannot be foregrounded and thus presented as a setting for the main clause; AP is neutral in this respect and therefore the only option in such cases.

All of these properties argue against treating AN simply as another relativization strategy in Kannada, whether externally headed (essentially those analyses that refer to the pronominal morpheme as the head) or, possibly, internally headed. While the purpose of relative clauses (RC) is to modify a nominal, the primary function of AN is to restructure the information flow in a particular way. This functional difference goes hand in hand with the differences in the internal structure of the AP vs. AN embedding. The restricted placement and the invariant morphological shape of the 'relativized' nominal make extremely doubtful the possibility of analyzing AN as an internally headed RC. Moreover, the presumed internal head is not always present, as is the case in the examples (3) and (12), and additional problems will be mentioned in section 5.

The externally headed alternative also faces complications beyond the fact that AN does not serve the same function as RCs. Specifically, AN and AP differ with respect to the permitted range of referents of the presumed head nominal. While there are no obvious constraints on the type of nouns AP can modify or complement, the distribution of pronouns in AN is restricted in two ways. First, it overwhelmingly favors human referents, indicated by the masculine and feminine pronouns. The grammatical neuter, which includes everything else, is rare in the singular and practically non-existent in the plural. The morphological structure of AN and the distribution of the pronominal elements in it are summarized in (17), using the nominative form as an example:

(17)　　[Tensed Stem – Nominalizer – Case]

malaguv	– *avan*	– *u*	masc. sg.
	– *aval*	– *u*	fem. sg.
	– *avar*	– *u*	masc./fem pl.
	– (*ud*	– *u*	neuter sg.)

Furthermore, the pronominal element is limited to the distal forms, marked by the initial *a-*,[4] while their proximal counterparts (*ivanu* '3SG:M', *ivalu* '3SG:F', *ivaru* '3PL:H'), or any other pronouns, for that matter, do not occur; we do not find forms such as **malaguv-ivanu* 'this one who is sleeping', **malaguv-ya:ru* 'which ones that are sleeping', etc. This suggests that the pronouns in AN are not independent heads of relative clauses but, rather, grammatical markers of a particular type of nominalization.

I propose to attribute all these apparent idiosyncracies to the special status of the AN constructions. The human vs. neuter asymmetry is at least reinforced, if not brought about directly, by the foregrounding function of AN: human referents are crosslinguistically much better candidates for topic-ness than non-human or inanimate ones, and this is also a robust distinction in the grammar of Kannada. The restriction to a single class of pronouns (distal, as it happens) can be, perhaps, tied to the resumptive function of AN: the pronominal element is not used here as a plain pronoun, but a special referential device whose primary function is to string two clauses together. As a result, the full range of referential possibilities normally associated with pronouns is not (and need not be) available in AN.

Finally, it is intriguing that the neuter form of AN is morphologically identical with G; compare the neuter form in the AN paradigm in (17) with the nominative form of G in (6). It could, of course, be just a coincidence. Some traditional descriptions say explicitly that the nominalizer in AN is a 'personal' pronoun only. I will show in the following section that this is not accurate, that the neuter singular form does occur in the resumptive construction as well, albeit not with equal frequency. And I will argue that it is precisely the referential properties of the neuter-based nominalizations together with the special referential status of the pronouns in AN that hold the key to our understanding of the relationships between AP, AN, and G.

5. Analysis

We start by observing that both AN and G occur in the so-called 'clefting' constructions (Sridhar 1990), which represent another commonly used pattern related to the resumptive construction discussed above. An example involving AN is in (18), which shows that the nominal referenced by the pronoun in AN appears post-verbally, in an unmarked form, and is interpreted as a contrastive focus:[5]

(18)   na:nu   pe:ṭeyalli   *no: ḍid-avan-u*   ra:ma
      1SG:NOM   market:LOC   see:PST-AN:SG:M-NOM   Rama:NOM
      '[As for] the one I saw in the market – that was RAMA.'

This pattern can be found in situations in which a specific piece of information from previous discourse is being denied or corrected. The sentence in (18), for example, could be an answer to questions such as 'Did you see the children in the market?', rather than 'Who did you see in the market?' The latter would trigger a simple focus-marking construction in which the focused constituent immediately precedes the verb. The sequence below shows such a neutral focus-marking construction in (19a) as distinct from an AN-based cleft construction in (19b), and a G-based cleft construction in (19c):

(19) a. pe:ṭege   ra:ma   ho:danu
      market:DAT   Rama:NOM   go:PST:3SG:M
      'RAMA went to the market.'
    b. pe:ṭege   ho:d-avan-u   ra:ma
      market:DAT   go:PST-AN:SG:M-NOM   Rama:NOM
      '[As for] the guy going to the market – that was RAMA.'
    c. pe:ṭege   ho:dad-d-u   ra:ma
      market:DAT   go:PST-G-NOM   Rama:NOM
      '[As for] the trip to the market – that was RAMA.'

Both AN and G in (19b) and (19c) are clearly used in the same structural position and serve the same discourse function – in both cases, the non-finite clause creates a setting for the nominal that follows. The crucial difference between them has nothing to do with whether the 'clefted' nominal is or is not the head of the embedded material (which would be an issue if we insisted on treating AN as a relative clause), since the gerundive variant in (19c) does not even raise this question. What matters is the referential vs. non-referential status of the non-

finite form: AN in (19b) makes an explicit reference to a participant in the event of going, while G in (19b) refers only to the event itself. But that in itself provides no basis for treating AN as a relative clause of any kind and, therefore, as a structure that would be formally distinct from G.

The role of referentiality in sorting out the relationship between G and AN comes out even more clearly in examples such as (20) below and (14b) above, here repeated as (21). Each sentence involves the neuter-form nominalization in a slightly different semantic context. In (20) two readings are available, while (21) allows only one of those two interpretations:

(20) se:bʰanṇu   me:jina   me:le   *iddad-d-annu* no:ḍidanu
     apple:NOM   table:GEN   on:top:of   be:PST-?-ACC   see:PST:3SG:M
     (i) 'That apple$_i$ on the table - he saw it$_i$.' (AN)
     (ii) '[There was an apple on the table]$_i$ – he saw it$_i$.' (G)

(21) se:bʰanṇu   me:jina   me:le   *iddad-d-annu* tindanu
     apple:NOM   table:GEN   on:top:of   be:PST-AN-ACC eat:PST:3SG:M
     'That apple$_i$ on the table - he ate it$_i$.'

The difference between (20) and (21) follows from the interaction between the semantics of the matrix verb and the referential possibilities of the neuter pronoun: anaphoric in (20-i) and non-anaphoric in (20-ii). The anaphoric use represents a AN reading of the nominalization, while the non-anaphoric one leads to a G reading, where the nominalization serves only to name the embedded event. This double reading is possible because the verb *no:ḍ-* 'see' takes either a nominal or a clausal complement for its second argument. Hence, *iddaddannu* can be interpreted either as AN, instantiating the nominal complement, or as G, serving as a *nomen actionis*. In contrast, the verb *tinnu-* 'eat' in (21) does not take a clausal complement and the G reading is, therefore, unavailable. The example in (20) thus provides evidence that it is not the nominalization *per se* that identifies the potential target for relativization – the selection of the target may depend on other factors, such as the semantics of the main verb (a similar observation has been made by Ohara 1994 about a partially comparable pattern in Japanese).

Based on examples of this kind, we can collapse AN and G and treat them as two uses of a single morphological structure with a split in the function of the nominalizing suffix, a distal pronoun. Depending on the nature of the nominalizer, this verbal noun can be associated with two different interpretations, provided that the subcategorization requirements of the matrix verb are compatible with one or the other reading. When the nominalizer is a personal pronoun, it results in an anaphoric AN construction. When it is the neuter pronoun, both anaphoric and non-anaphoric interpretations are available. The result is a potential ambiguity between AN and G readings with verbs that permit complements whose referent may be either a proposition (required for the G reading) or an entity (AN reading). The ambiguity is more readily resolved in favor of the non-anaphoric option (G), for pragmatic reasons: the grammatically neuter referents are not good candidates for the resumptive anaphoric function served by AN. This asymmetry, then, provides motivation for reanalyzing the non-anaphoric function as a 'pure'

nominalizer, leading to the apparent split between ANs and Gs as two distinct categories: a referential verbal noun vs. a plain *nomen actionis*.

This analysis also helps answer the question posed at the beginning, namely why the verbal noun often fails to occur in the genitive form in the adnominal position and why AP is used instead. The failure is illustrated in (5c) above, in contrast to the successful use in (8b), here repeated as (22a, b), respectively: the verbal noun *baruvudara* 'of coming' as a complement of *varege* 'to the point of' is permitted in (22b), while the verbal noun *biṭṭuho:guvudara* 'of leaving' as a complement of *vis^haya* 'knowledge of' in (22a) is not.

(22) a. *biṭṭuhooguv-a/*biṭṭuhooguv-ud-ara*   varege   baa
     leave:NPST-AP/*-G-GEN   to:point:of   come:IMP
     'Come to the place of departure.'

   b. ra:ma illige   *baruv-ud-ara*   vis^haya   nanage   ma:tana:ḍabe:ku
     Rama here   come:NPST-G-GEN   about   1SG:DAT   talk:MOD
     '[Rama's coming here]$_i$; I want to talk about it$_i$.'

The ungrammaticality of the verbal noun in (22a) stems from the fact that it cannot be interpreted anaphorically, as an instance of AN use: the verbal noun is just a *nomen actionis*, without any referential entity that could be the target of the resumptively used nominalizer under the anaphoric reading. However, if the embedded non-finite clause headed by the quasi-nominal simply identifies an event, without any anaphoric relationship to one of its own constituents, this use of the verbal noun becomes indistinguishable from the function served by AP, which is a highly productive form fully dedicated to putting a verbal element into a noun-governed position. The competition is thus easily resolved in favor of AP over G on functional grounds, even though a genitive G would be expected paradigmatically. This result may be further reinforced by the fact that the nominalization is formally the same for both the G and AN readings and the gap in the G paradigm thus helps resolve the functional indeterminacy in that only the anaphoric reading (AN) is permitted.

Against this functional background, we can now summarize the case marking patterns as follows. The verbal noun is inflected for all cases when used anaphorically (AN reading), as would be expected, while it is not inflected for the genitive when used non-anaphorically (G reading). In verb-governed positions, this gap follows from the case-marking patterns of verb-governed complements and is independent of the verbal noun itself. Adnominally, however, it is forced out by a competing form, namely AP. All these relationships are summarized in Diagram 1. The downward arrows indicate (very broadly identified) functional domains associated with the three forms in question, the upward arrows represent the mapping between each of the forms and the function(s) they serve. The symbol X in the morphological template for the participles stands for a class of participial suffixes, only one of which is the relative suffix –*a* that yields AP.

Diagram 1. Network of forms and functions representing AN, G, and AP.

NOMINALIZATION   PARTICIPLES

[stem - nominalizer - ...]   [stem - X]

'pers. pro'   neuter sg.   adnominal   adverbal
(masc., fem.)

anaphoric   non-anaphoric

functions

     verb, noun   verb complement   noun compl.   modifier
     compl.

forms   all cases   all cases   *Genitive
                governed by V
                (hence *Genitive)

     AN   G   AP
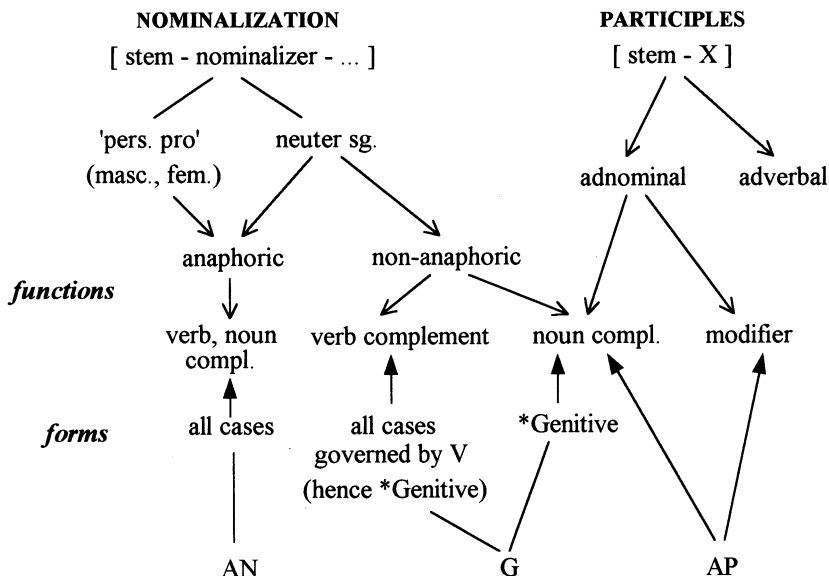
6. Conclusions and consequences

The goal of this paper has been to demonstrate that a closer examination of morphological, semantic, and pragmatic properties of partially overlapping expressions can yield a network of distinct grammatical patterns that help explain an otherwise puzzling distribution displayed by one of the expressions. The overlapping functions of two expressions in a particular environment (G and AP) lead to a paradigmatic gap in one of them, while the overlapping form in two other expressions (G and AN) may further strengthen the gap. Thus I conclude that the gap results from two mutually reinforcing pressures that work toward maintaining a relatively transparent network of function-form relationships.

On a more general level, the proposed analysis contributes to our understanding of the relationship between relativization and nominalization as two distinct grammatical devices. As the Kannada material shows, the distinctness can be superficially obscured by the fact that they share certain morphosyntactic properties: both represent non-finite embedding strategies and both revolve around a nominal complement of the matrix verb. However, they part company along a number of other, more specific criteria, both functional and formal. Overall, then, the analysis highlights the interrelatedness of grammatical patterns, suggesting that to understand one pattern means to also understand certain other patterns. The challenge is to determine precisely those properties and

relationships that are at the heart of the network, and I tried to address this question in this paper. What remains to be done is to work out the issues involved in representing such networks.

Notes

1 Unless marked otherwise, the data used in this paper were obtained by the author through extensive elicitation from a single female speaker of one of the South-Central dialects of Kannada.
2 LOC 'locative', NOM 'nominative', DAT 'dative', ACC 'accusative', GEN 'genitive', ABL 'ablative', G 'gerund', AN 'anaphoric nominalizer', AP 'adnominal participle', PST 'past', NPST 'non-past', PPF 'present perfect', FPF 'future perfect', PPL 'past participle', IMP 'imperative', MOD 'modal', NEG 'negation', SG 'singular', PL 'plural', M 'masc.', F 'feminine', N 'neuter', H 'human'.
3 Kannada has a number of expressions of somewhat unclear grammatical status, such as *vis^haya* 'knowledge', *bagge* 'about', *me:le* 'on to of', *o*ʃ*age* 'at the time of', etc., that behave like nouns in that they govern a genitive-marked nominal but functionally and morphologically resemble adpositions (they are indeclinable expression of various adverbial relations). For the purposes of this paper I will assume that structurally they can be analyzed as nominals.
4 The *udu* form in the neuter is old and more common in the literary style, while *adu* is an innovation of the colloquial register (Spencer 1914). My consultant consistently uses the *adu* variant as the independent distal pronoun and the *udu* variant in the nominalizations.
5 It may very well be that contrastiveness is only one in a range of functions served by the clefting; relevant factors may also include afterthought, newsworthiness in the sense discussed in a number of papers in Payne1992, and perhaps others. The exact nature of the discourse conditions remains to be worked out.

References

Andronov, Mikhail (1979). 'Verbals in Dravidian: a comparative study.' IJDL 8, 52-70.

Ohara, Kyoko Hirose (1994). 'An event-reporting relative construction in Japanese.' BLS 20, 260-272.

Nadkarni, M.V. (1972). 'Sentential complements on nouns in Kannada.' *Indian Linguistics* 33, 1-23.

Payne, Doris L., ed. (1992). *Pragmatics of Word Order Flexibility*. Amsterdam: John Benjamins.

Shiffman, Harold (1983). *A Reference Grammar of Spoken Kannada*. Seattle: Washington University Press.

Spencer, Harold (1914). *A Kanarese Grammar*. Mysore: Wesleyan Mission Press.

Sridhar, S.N. (1990). *Kannada*. London: Routledge.

Steever, Sanford B. (1988). *The Serial Verb Formation in the Dravidian Languages*. MLBD Series in Linguistics 4. Delhi: Motilal Banarsidass.

Steever, Sanford B. (1998). *The Dravidian Languages*. London: Routledge.

The phonology of the past tense in Tamil[*]

CAROLINE R. WILTSHIRE

University of Florida, Gainesville

1. Introduction

Tamil is a Dravidian language spoken in Sri Lanka and Tamil Nadu, India. Verbs in Tamil are traditionally divided into two major categories, strong vs. weak, with seven classes of tense markers (Arden 1942); this classification is still used in teaching Tamil (Lindholm and Paramasivam 1980). The two major classes are based on the form of the present tense marker, a single -*k*- in weak classes and double -*kk*- in strong. Five forms are used for the past tense: -*t*-, -*nt*-, -*in*-, doubling, and -*tt*-.

(1)

	Class	Root	Past	Present	Gloss
Weak:	1a,b	sey-, koḷ	-t-	-kiɾ-	'do', 'kill'
	2	viḻ-	-nt-	-kiɾ-	'fall'
	3	tiɾump-	-in-	-kiɾ-	'return' (intrans.)
	4	viṭ-	double	-kiɾ-	'leave'
	5a	en-	-t-	-kiɾ-	'say'
Strong:	5b	keeḻ-	-t-	-kkiɾ-	'ask'
	6	paaɾ-	-tt-	-kkiɾ-	'look'
	7	naṭa-	-nt-	-kkiɾ-	'walk'

Two forms of the past appear in both weak and strong classes (5a/5b, 2/7), and Class 1 is subdivided based on the result when the past is added.[1] The five forms undergo further variation in the phonology, resulting in at least nine phonetic variants (2).[2]

(2)

	past	imper.	past+3fsg	gloss	Class	Past UR
a.	[in]	tiɾumbu	tiɾumbinaa	'she returned'	3b.	/in/
b.	[nd]	naḍa	naḍandaa	'she walked'	7	/nt/
c.	[ɳḍ]	koḷu	koɳḍaa	'she had'	1b	/nt/
d.	[ɲj]	eɾi	eɾiɲjaa	'she burned'	2	/nt/
e.	[ð]	aḻu	aḻuðaa	'she cried'	1a	/t/
f.	[ṭ]	viḍu	viṭṭaa	'she left'	4	/t/
g.	[tt]	koḍu	koḍuttaa	'she gave'	6	/t/
h.	[cc]	ciɾi	ciɾiccaa	'she laughed'	6	/t/
i.	[ṭṭ]	keeḷu	keeṭṭaa	'she asked'	5b	/t/

Building on the insights of Christdas (1988) and Vasanthakumari (1989), I argue here

that only three forms of past tense markers are required to account for the phonology of Tamil past tense formation. By using output constraints, this account brings motivation to the system of alternations and captures generalizations missed by Christdas (1988). The formulation of constraints also does away with some of the arbitrary specifications required in a rule-based account, such as the direction of spreading, the choice of consonants to be deleted, and the ordering of rules. Furthermore, although Christdas required that the past tense markers be added at two different levels in her Lexical Phonology model, interleaved with level-specific rules, I show that a single level is sufficient in a constraint-based framework such as Optimality Theory (Prince and Smolensky 1993).

2. Analysis of past tense realization in OT

The analysis within OT relies on a set of cross-linguistically motivated constraints (ˈ2.1), ranked in a language-specific order for Tamil (ˈ2.2). The input representations used here for the past tense are indicated on the right in (2). The choice of affix is determined by the morphological class of the root: /t/ for Classes 1a-b, 4, 5a-b, and 6, /nt/ for 2 and 7, and /in/ for 3. The realization of /in/ is straightforwardly [in], while the realizations of input /t/ and /nt/ are output [ð,tt,cc,ʈʈ,nd,ɳɖ,ɲj], with assimilation or deletion of stem-final sonorants. Rather than a series of rules, the surface-based constraints active in Tamil phonology combined with universal correspondence constraints determine how output varies from input.

2.1 Constraints

The constraints involved in the realization of the past tense have been previously motivated both cross-linguistically and in accounts of Tamil noun phonology (Beckman 1998, Kaun 1998, Wiltshire 1998a). First, assuming sonorant geminates are moraic, Tamil obeys the cross-linguistic tendency to avoid trimoraic syllables, *TRI-μ. Syllables may thus have long vowels (3a) or short vowels plus a sonorant coda (3b), but long vowels followed by geminate sonorants are not allowed (3c).

(3) a. single sonorant after long V: [kaalɯ]   'leg'
   b. moraic sonorant after short V: [kallɯ]   'stone'
   c. *moraic sonorants after long V: *[kaallɯ]
   d. linked nasal+obstruent after long V: [paambu]   'snake'
   e. geminate obstruents after long V: [kaattɯ]   'wind' (noun)

Long vowels may be followed by a linked nasal plus stop (3d) or geminate obstruents (3e). I will stipulate that laterals, rhotics, and glides in coda position are moraic, but nasals and obstruents are not (see Kaun 1999 for an alternative account).

The second cross-linguistically common constraint is NoCODA$_{PL}$, which is violated by a coda consonant with a place of articulation distinct from that of the

following onset. While this constraint is violable in the first syllable of nouns (see Beckman 1998 for a positional-faithfulness account), such violations are not tolerated in verb forms. Both assimilation (4a) and epenthesis (4b) are motivated by this constraint (Wiltshire 1998a). Epenthesis is also motivated by the constraint *COMPLEX, first proposed by Prince and Smolensky (1993), which bars clusters of consonants in onsets and codas, as in (4b).

(4) a. /maɾam+kaḷ/   'tree, pl'   [maɾaŋga]   'trees'
     b. /payyan+kk/   'boy, dat'   [payyanɯkkɯ]   'to the boy'

The three general phonotactic constraints discussed so far are given in (5) below.

(5) a. *TRI-μ: No trimoraic syllables. (Christdas 1988:259, Kaun 1998)
     b. NOCODA$_{PL}$:   C-Place in onset position only. (Wiltshire 1998a)
     c. *COMPLEX: No C clusters in onset or coda (Prince and Smolensky 1993)

In order to discuss the remaining constraints, I provide the phonemic inventory of Tamil in Table (6), which also illustrates the (fairly traditional) transcriptions I use here. Tamil has a five vowel system, and also distinguishes long vowels from short.

(6) Phonemic inventory of Tamil

Vs	Front		Central		Back		
High	i ii					u uu	
Mid	e ee					o oo	
Low			a aa				
Cs Place6 Manner9	Labial	Dental/	Alveolar	Retroflex	(Al)Pal	Velar	
Stop	p	t	t̲	ʈ	c	k	
Nasal	m		n	ɳ	ɲ		
Fricative			s	ʂ	ʃ		
Lateral glide			l	ɭ			
Tap			ɾ				
Glide	ʋ			ɻ	y		

Many of the consonants in the inventory can be single or geminate. The phonemic inventory of Tamil shows three sub-places of articulation within the coronal region.

For most consonants, the dental-alveolar contrast is not distinctive, though in Christdas's dialect dental and alveolar stops do contrast.

All dialects of Tamil distinguish between dental, palatal, and retroflex obstruents and nasals, but dental stops become palatal after [i,e,y] and retroflex after retroflex consonants. Following Gnanadesikan (1994) and Hume (1994), this can be captured by analyzing front vowels as having a coronal node with a laminal articulation. Dental coronals are the most unmarked, and hence have no subnode; palatals have a laminal subnode like front vowels, and retroflexes have an apical node.

(7) Front Vs   Dentals   Palatals   Retroflexes
       V   C   C   C
       | | | |
      Cor   Cor   Cor   Cor
       | | |
      Lam   Lam   Apical

In palatalization, a dental takes on the laminal articulation of an adjacent vowel or consonant and becomes a palatal. Retroflexes resist due to their apical specification. In retroflexion, a dental takes on the apical node of an adjacent consonant. Palatals resist due to their laminal specification.

To force the spreading of coronal subplaces, I have formulated the constraint in (8a), which states that coronal subnodes are realized on post-vocalic consonants. The motivation for this constraint is acoustic. While most place features are most salient during the release of a consonant onto the following sonorant, some, like retroflexion, are unusual in that their cues are most salient during the transition from a vowel to a consonant (Steriade 1995). Hence the constraint in (8a) states that subcoronal features, laminal and apical, appear in post-vocalic position.

(8) a. SUBCOR(V_):   Coronal subplaces are realized on post-vocalic Cs.
     b. HAVEPLACE   Consonants have a place of articulation.

The phonotactic (8b), forces consonants to take on some place of articulation (c.f. FILL^PLACE from Prince and Smolensky 1993:181). This supports assimilation in positions in which independent place features are prevented, such as in the coda, and in consonants which do not have place features in input.

The inventory in (6) shows only voiceless stops, as voicing is not distinctive in native Tamil words. Generally in Tamil, post-nasal obstruents are voiced. As observed by Pater (1995), many languages do not allow sequences of nasals followed by voiceless obstruents; such underlying clusters surface with a voiced obstruent or without the nasal. Pater proposes a cross-linguistically justified constraint:

(9) *NC$_{vl}$:   No nasal-voiceless obstruent sequences (Pater to appear)

In Tamil, input sequences of nasal plus voiceless obstruent surface with voiced obstruents due to the high ranking of this constraint.[3]

The above constraints are formulated in terms of permissible surface structures. The second kind of constraints required are correspondence constraints (McCarthy and Prince 1995), which evaluate how input and output differ from each other. MAX-IO constraints force underlying segments of input to surface in output; deletion of a consonant violates (10a), for example. DEP-IO constraints regulate whether output matches input; an epenthetic vowel in the output is a violation of (10b). Finally, to limit assimilation, the IDENT family of constraints requires a segment in output to have the same features as its corresponding segment in input (10c).[4]

(10) Correspondence Constraints (McCarthy & Prince 1995)
    a. MAX-IO(C):   Input consonant has a correspondent in the output.
    b. DEP-IO(V):   Output vowel has a correspondent in the input.
    c. IDENT-IO(X):   Output segments have the same specification for the feature [X] as do their correspondents in the input.

The relationship between phonotactics and correspondence constraints is one of language-specific ranking. To account for surface variation in the past tense marker in Tamil, we will find a single ranking which matches inputs and outputs. The phonotactics important throughout Tamil will rank high, showing that the alternations in the past tense markers reflect general properties of Tamil rather than requiring arbitrary rules.

2.2 Rankings

The realizations of the past tense forms relative to the input involve voicing, assimilation, and gemination, while stem-final sonorants are subject to deletion. The ranking of the constraints is shown to be general, but requires a few statements about the assumptions of the underlying forms of stems. The stem is determined based on its singular imperative, but if the final vowel of that form is /u/, it is assumed to be epenthetic. In some cases, following Christdas (1988), the UR of the stem will be represented with a final empty C-slot; such cases are motivated in §2.2.5.

2.2.1 Bleeding of epenthesis: /-in/

Assuming that the final vowel /u/ is not part of the root but rather appears word-finally for syllabification (Wiltshire 1998a), the lack of epenthesis before the vowel-initial past suffix /-in/ is easily explained. An epenthetic vowel appears when *COMPLEX would otherwise be violated, as in the imperative in Tableau 1; this motivates the ranking of *COMPLEX above DEP(V). In forms in which the past tense suffix is /-in/, as in (2a), an epenthetic vowel is superfluous, as *COMPLEX is in no

danger of being violated in the faithful surface candidate; this is shown in the past tense form in Tableau 2.

Tableau 1 Input: /tiɾump/ = [tiɾumbu] 'return!'

Candidates	*COMPLEX	DEP(V)
~~tiɾumb~~	*!	
√ tiɾumbu		*

Tableau 2 Input: /tiɾump + in + aaɭ/⁵ = [tiɾumbinaa] 'she returned'

Candidates	*COMPLEX	DEP(V)
~~tiɾumbuinaa~~		*!
√ tiɾumbinaa		

2.2.2 Post-nasal voicing: /-nt/

The form of the past tense in /nt/ shows a voiced obstruent in all its phonetic realizations, as in (2b,c,d). This is attributed to the high ranking of the *NC$_{vl}$ constraint from (9), whose force is seen throughout Tamil, above the correspondence constraint IDENT([voice]). In the forms of class 7, in which /nt/ surfaces unchanged except as to voicing, this ranking correctly provides the past tense form as [nd]:

Tableau 3 Input: /naʈa + nt + aaɭ/ = [naɖandaa] 'she walked'

Candidates	*NC$_{vl}$	IDENT([voice])
~~naɖantaa~~	*!	*
√ naɖandaa		**

2.2.3 Assimilation: /-nt/, /-t/

The /-nt/ and /-t/ forms of the past tense marker are also subject to two forms of assimilation: palatalization to give [ɲj] and [cc] in (2d,h), and retroflexion to give [ɳɖ] and [ʈʈ] in (2c,f,i). In all cases, assimilation extends from a verb-final segment rightward onto the suffix. In some cases, assimilation is accompanied by the loss of the triggering segment, which requires ordering in a rule based account; the account here is compatible with having no ordering between the two (§2.2.4).

Retroflexion comes from a retroflex stem-final consonant. In the simplest cases, retroflex-final verb stems, like *viʈ* 'to leave', surface with a geminate [ʈʈ] in the past tense. This is due to ranking the constraint NOCODA$_{PL}$ above IDENT([subcor]). Both

candidates in Tableau 4 satisfy SUBCOR(V_), so its ranking is not yet determined; NOCODA_PL alone guarantees the assimilation.

Tableau 4 Input: /vit̪ + t + aal̪/ = [vit̪t̪aa] 'she left'

Candidates	NOCODA_PL	IDENT([subcor])
vit̪taa	*!	
√ vit̪t̪aa		*

Palatalization shows that SUBCOR(V_) ranks above IDENT([subcor]). When the /-nt/ past is added to a verb ending in a front vowel, palatalization extends through the cluster due to the combined effects of SUBCOR(V_) and NOCODA_PL. High ranking *NC_vl continues to force the voicing of post-nasal obstruents. Hence the result is [ɲj].

Tableau 5 Input: /eɾi + nt + aal̪/ = [eɾiɲjaa] 'she burned'

Candidates	SUBCOR(V_)	NOCODA_PL	*NC_vl	IDENT ([subcor])	IDENT ([voice])
eɾintaa	*!		*!		
eɾindaa	*!				*
eɾiɳdaa		*!		*	*
eɾiɲcaa			*!	**	
√ eɾiɲjaa				**	*

2.2.4 Loss of sonorant with assimilation: /-nt/

Stem-final sonorants are lost before /-nt/ due to two high ranking constraints, *TRI-μ and *COMPLEX, which limit syllable size. For a verb such as *aal̪* 'rule', adding the past /-nt-/ results in output *aaɳḍaa* 'she ruled'.

Tableau 6 Input: /aal̪ + nt + aal̪/ = [aaɳḍaa] 'she ruled'

Candidates	SUBCOR(V_)	*TRI-μ	MAX(C)	IDENT([subcor])
aalntaa	*!	*!		
aaldaa		*!	*	*
aadaa			**!	*
√ aaɳḍaa			*	**

The retroflex place is preserved and spread into the past marker due to the interaction between SUBCOR(V_), which requires the retroflex place feature to survive on the coda consonant, and the NOCODA$_{PL}$, which requires codas to share place features with the following onset. The loss of the lateral is motivated by the constraint that syllables may not be trimoraic, *TRI-μ; the lateral is lost rather than some part of the suffix since the loss of a nasal or obstruent will not satisfy *TRI-μ.

For verbs like *koḷ* 'have' or *sey* 'do' with short vowels, the addition of the past tense /-nt-/ also results in surface forms without a final sonorant: *koṇḍaa* 'she had' and *seɲjaa* 'she did', although *TRI-μ is not at stake. The loss of the final sonorant is due to the phonotactic against complex clusters in onset or coda, *COMPLEX. Furthermore, that laterals and glides are lost rather than part of the suffix is due to the higher ranking of a specific correspondence constraint that preserves nasals and obstruents, MAX(C-[-cont]), above the general constraint MAX(C). A similar ranking of MAX(C) sub-constraints has been motivated for IruLa (Wiltshire 1994).

Tableau 7 Input: /koḷ + nt + aaḷ/ = [koṇḍaa] 'she had'

Candidates	SUBCOR (V_)	*COMPLEX	MAX (C-[-cont])	MAX(C)	IDENT ([subcor])
kondaa	*!			*	
koḷndaa		*!			
koḷḍaa			*!	*	*
√ koṇḍaa				*	**

Tableau 8 Input: /sey + nt + aaḷ/ = [seɲjaa] 'she did'

Candidates	SUBCOR (V_)	*COMPLEX	MAX (C-[-cont])	MAX(C)	IDENT ([subcor])
sendaa	*!			*	
seyndaa		*!			
sey.jaa			*!	*	*
√ seɲ.jaa				*	**

These surface forms of the past tense verbs appear to have undergone right-spreading assimilation followed by consonant deletion. Rather than order such rules, the OT account chooses the optimal output candidate by simultaneously evaluating the

constraint ranking in Tamil on a set of candidate outputs.

2.2.5 Gemination: /-t/

In some cases, as in Class 5b, assimilation results in a geminate obstruent (2i). While the stem-final sonorant is not totally lost, in order to satisfy *TRI-μ it can no longer retain the feature [lat] from input, hence violating IDENT([lat]) in Tableau (9):

Tableau 9 Input: /keeḷ + t + aaḷ/ = [keeṭṭaa] 'she asked'

Candidates	NOCODA_PL	*TRI-μ	SUBCOR(V_)	MAX(C)	IDENT([lat])
keeḷtaa	*!	*!			
keeḷḍaa		*!			
keeldaa		*!	*!		
keeḍa				*!	*
√ keeṭṭaa					**

Similarly, stems ending in /ɾ/, such as *paaɾ* 'look', which take the past /-t/, appear without the final rhotic, which has assimilated due to *TRI-μ to become a geminated /tt/. As neither /ɾ/ nor /t/ has a subcoronal place, SUBCOR(V_) is not at issue here:

Tableau 10 Input: /paaɾC + t + aaḷ/ = [paaṭṭaa] 'she saw'

Candidates	NOCODA_PL	*TRI-μ	SUBCOR(V_)	MAX(C)	IDENT([ɾ])
paaɾtaa	*!	*!			
paaɾtaa		*!			
paataa				*!	*
√ paaṭṭaa					*

For other cases showing geminates in the past (2g,h), I use the inputs forms argued for in Christdas. She distinguishes verbs that end in a specified consonant or vowel from verbs that end in an unspecified consonantal slot, here represented with C. This allows us to account for the geminated tense suffixes that appear on these verbs in all three tenses. Gemination of features to an empty C-slot requires that constraint (8b), HAVEPLACE, outranks the constraint requiring identity of output features, IDENT([place]). Two cases are shown below. In the first, Tableau (11), only the major place [cor] appears on the output consonant corresponding to the input C-

slot. In the second, Tableau (12), the subplace from the adjacent front vowel spreads to both the C-slot and the past tense marker, due to the high ranking of SUBCOR(V_) and NOCODA_PL.

Tableau 11 Input: /koṭuC + t + aal/ = [koḍuttaa] 'she gave'

Candidates	HAVEPLACE	IDENT([place])
koṭuCtaa	*!	
√ koḍuttaa		*

Tableau 12 Input: /ciriC + t + aal/ = [ciriccaa] 'she laughed'

Candidates	SUBCOR(V_)	HAVEPLACE	NOCODA_PL	IDENT([place])
cirittaa	*!			*
ciriCtaa	*!	*!		*
cirictaa			*!	*
√ ciriccaa				**

2.2.6 Summary of rankings

The rankings used for the alternations in the past tense forms are given in (11); I have separated three sets of rankings since many of the constraints cannot be directly ranked relative to one another. (11a) forces the change in voicing after a nasal. (11b) shows that three constraints can force a mismatch between underlying and surface place features: SUBCOR(V_), HAVEPLACE and NOCODA_PL.

(11) a. *NC_vl   b. {SUBCOR(V_), HAVEPLACE, NOCODA_PL}
           | \ | /
      MAX-IO([voice])   {IDENT([place]), IDENT([subcor])}

      c.   {*COMPLEX, MAX([-cont]), *TRI-μ }
               \   \   /
                     MAX(C)
                        |
             { IDENT([lat]), IDENT([r]) }

(11c) gives the constraints that force deletion or the mismatch of manner features: *COMPLEX, MAX([-cont]), and *TRI-μ. Each of the high-ranked constraints is motivated both cross-linguistically and in Tamil generally. The use of constraints in

analyzing the past-tense alternations reveals that they are not arbitrary and isolated, but fit naturally into the system of Tamil phonology.

3. Comparison with Alternatives

In the analysis presented here, rule ordering and level ordering are unnecessary. In Christdas (1988), tense markers are added at two different levels in a Lexical Phonology account. In Level 1, the past tense is added for all strong verbs and for weak verbs which take /-t/ and /-nt/. In Level 2, the past tense for the remaining weak verbs, /-n/, is added, with a rule of epenthesis supplying the vowel before it. In Christdas' analysis, verb inflection precedes some of the more productive derivational morphology, and strong verbs are treated as relatively more irregular than weak verbs; both are questionable morphologically but necessary for her phonological system of rules to work. She also requires ordering relationships, so that, for example, coronal assimilation between a stem and past tense marker bleeds epenthesis. Such rule orderings are avoided here by prioritizing the constraints.

Furthermore, using rules provides the formal power to write similar rules that are unmotivated cross-linguistically, such as adding a vowel to vowel clusters rather than to unsyllabifiable consonant clusters. Constraints, by contrast, are general and cross-linguistically motivated, and reranking them does not produce impossible languages. Furthermore, constraints capture the motivation behind alternations, and as these motivations are general throughout Tamil, the same constraints should account for further facts of Tamil and related languages, a project I leave for future research.

4. Conclusions

The analysis has reduced the number of classes of past tense markers from five to three, and the number of levels to input and output, without intermediate stages. Output forms have been motivated by phonotactics, so that the account is both simpler and more general than previous analyses. The constraints invoked here are also found highly ranked in related languages, and allow us to state the similarities between them. For example, most Dravidian languages have some form of a coda-place constraint motivating assimilation and/or deletion (Telugu, IruLa, and Malayalam in Wiltshire (1994, 1998b), Mohanan (1982, 1998)). Dentals commonly assimilate to retroflexes and palatals, favoring preservation of the subplaces of coronal over the unmarked dental place, though this may be more general than in Dravidian alone (Gnanadesikan 1994). Furthermore, IruLa similarly preserves obstruents and nasals at the expense of glides and laterals (Wiltshire 1994). The case study of the past tense shows that its alternations are motivated by constraints relevant for Tamil, for Dravidian, and more generally.

Notes

53

¹ Some classes can be described in phonological terms (Lindholm and Paramasivam 1980): 1b=lat.
-final weak, 3=non-minimal [u]-final, 4=minimal [ṭu]/[ru]-final, 5a =nas.-final, 5b=lat.-final strong,
7=all [a]-final plus *iru* 'to be'. Two others, 2 (weak) and 6 (strong), act as default classes. Correlations
between these observations and the choice of past marker are beyond the scope of this paper.
² See (6) for an indication of the transcription used here. Although the palatal stop symbol /c/ is
used, it traditionally indicates an alveopalatal affricate in Tamil.
³ A similar analysis for allophonic intervocalic lenition, a general process of Tamil seen in (2e),
is omitted for reasons of space, but see Kaun (1998), who uses a LENITE constraint.
⁴ Since the correspondence constraints here compare input and output, I omit '-IO' to save space.
⁵ The final /ḷ/ of the 3rd f.sg. suffix appears if a vowel-initial suffix follows, such as the question
marker -*aa*. Its deletion word-finally is not explicitly analyzed here.

References

Arden, A.H. (1942). *A Progressive Grammar of Common Tamil*, Fifth edition, Madras, India: The Christian Literature Society.
Beckman, J. (1998). Positional Faithfulness. PhD dissertation, UMass, Amherst.
Christdas, P. (1988). The Phonology and Morphology of Tamil. PhD diss., Cornell.
Gnanadesikan, A. (1994). The Geometry of Coronal Articulations. in M. Gonzàlez, (ed.), *NELS* 24, 125-139.
Hume, E. (1992). Front Vowels, Coronal Consonants, and their Interaction in Nonlinear Phonology, PhD dissertation, Cornell. Published 1994, NY: Garland.
Kaun, A. (1998). Input Constraints in Tamil. Paper presented at CLS 34, Chicago.
Kaun, A. (1999). Using English to Understand Tamil: Sonorants, Obstruents, and Weight. Ms., Yale University.
Lindholm, J. and K. Paramasivam. (1980). *A basic Tamil reader and grammar*. University of Chicago South Asia Language and Area Center.
McCarthy, J. and A. Prince (1995). Faithfulness and Reduplicative Identity. In Beckman et al. (eds) *Papers in Optimality Theory*, Amherst: GLSA, 249-384.
Mohanan, K.P. (1982). The theory of lexical phonology. PhD dissertation, MIT.
Mohanan, K.P. (1998). On the Bases of Optimality Theory, Paper presented at SALA XIX, York University.
Pater, J. (to appear) Austronesian Nasal Substitution and other NC effects. In Van der Hulst, Kager, and Zonneverld (eds)., *The Prosody Morphology Interface*. Cambridge: Cambridge University Press.
Prince, A. and P. Smolensky. (1993). *Optimality theory*. TechRep #2, RuCCS, NJ.
Steriade, D. (1995). Positional Neutralization, Ms., UCLA.
Wiltshire, C. (1998a). Extending ALIGN constraints to new domains. *Linguistics* 36-3: 423-467.
Wiltshire, C. (1998b). Consonantal Place and Voice Features: Interactions in Telugu sandhi, presented at SALA XIX, York University.
Wiltshire, C. (1994). Feature constraints in IṛuLa. Beals et al.(eds.) *CLS* 30:428-442.
Vasanthakumari, T. (1989). *Generative Phonology of Tamil*. Mittal: Delhi.

Analyzing contact-induced phenomena in Karaim

ÉVA ÁGNES CSATÓ
Uppsala University, Sweden

1. Peripheral Turkic languages

Deviant developments are particularly strong in the periphery of the Turkic-speaking world. Turkic languages spoken in so-called language islands, separated from the Common Turkic linguistic area, have often converged in a fascinating way with dominating non-Turkic languages. Most of the peripheral Turkic languages exhibit highly interesting examples of contact phenomena, new lexical elements, syntactic structures, and articulatory habits copied from non-Turkic contact languages.

2. Code-copying

The Model of Code-Copying elaborated by Johanson (1992, 1993, 1996) serves as a typological framework in which different types of copying processes are defined. The term copying is introduced to replace traditional terms such as borrowing and code-switching. The model distinguishes between global, selective and mixed copies of material, as well as combinatory, semantic and frequential features. One of the basic tenets of the model is that a copy is never identical with the original and that copying processes always lead to innovations and change the copying code.

3. The case of the Karaim language

Here I will present examples illustrating different types of copying in the Lithuanian dialect of the Karaim language. Karaim belongs to the Kipchak branch of the Turkic language family. Its two living dialects, the Lithuanian and the Ukrainian, manifest a number of typological properties shared with cognate Turkic languages. They have, however, undergone considerable changes induced by language contact. The adaptation of new structural and lexical properties has led to a significant convergence with dominant areal linguistic types. This paper focuses on the nature of linguistic processes involved in such typological changes in general. Both dialects are highly endangered. The Lithuanian dialect is spoken today by about 40 speakers and the Ukrainian dialect by six speakers altogether, of which only two have full-fledged competence.

4. Frame-changing processes

The theoretical issue to be studied is how copying may change the structure of the copying language. Karaim data provide arguments to claim that the distinction between frame-changing and non-frame-changing processes is not a crucial one. In principle, each new copy changes the frame. I take the position that contact-induced language change occurs at all levels of linguistic structure. Continuous copying leads to increasing similarity of the codes involved. More equivalence positions for insertion of new copies are created. By incorporating new items and structures, the basic code is permanently reshaped to form the basis for new frame-changing developments.

Different types of changes are triggered by the process of copying. Global copies of lexical material may introduce new syntactic patterns, as illustrated by

adpositional and case-assigning phenomena in Karaim. Like other Turkic languages, Karaim is a postpositional language. The copy of the Slavonic preposition *okolo* meaning here 'about' introduced a new syntactic structure in which the adposition precedes the nominal complement. The case-assigning properties of the preposition, i.e. its combinatorial features, were also copied. Thus, the preposition ascribes the Karaim genitive to the noun *afta* 'week', just as the Slavonic preposition also requires the genitive.

(1)　*Karaim*

Okolo	b'ir	aftanïn	Troxta	kalïm.
about.PREP	a	week:GEN	Trakai:LOC	stay:AOR:1SG

'I will stay about a week in Trakai.'

Frequential copying may change the functional load of some constructions and thus lead to the loss of certain 'typical' constructions, as the two different types of Karaim 'have'-constructions illustrate. Both constructions are grammatical; the frequency of construction (2b) has increased as a result of contact with Slavonic, in which this is the usual 'have' construction. (2a) illustrates the typical Turkic 'have' construction.

(2)　*Karaim*

a.
m'en'im	ed'i	ek'i	karïndašïm
I:GEN	be:PST	two	brother:POSS1SG
N:GEN	COPULA	NP:POSS	

'I had two brothers.'

b.
Troxta	bart	b'iz'd'à	uŋlu	yùv.
Trakai:LOC	existing:3	we:LOC	big	house
	COPULA	N:LOC	NP	

'We have a big house in Trakai.'

Copied word order patterns have changed the markedness patterns of certain Karaim constructions and consequently led to basic typological changes. Today Karaim is an SVO language. The following table compares the Karaim word order features with those in the dominating languages of the area: Lithuanian, Polish and Russian. The cognate Turkic languages Karachay-Balkar and Turkish have an SOV order.

(3)　*Basic word order in Karaim, its cognates (Turkish and Karachay-Balkar) and the dominating contact languages (Lithuanian, Polish and Russian)*

Karaim	SVO	Postp / (Prep)	NG / GN	AN / (NA)
Turkish and Karachay-Balkar	SOV	Postp	GN	AN
Lithuanian	SVO	Prep	GN	AN
Russian	SVO (free variation)	Prep	NG	AN
Polish	SVO (free variation)	Prep	NG	AN / NA

The hypothesis I wish to present here is that operators are especially sensitive to copying processes. I will apply the term *operator* to phonological and grammatical elements which bear a scope relation to some part of a phonological form or a

syntactic structure, such as a quantifier or a negative. *Scope* is taken to be a part of a linguistic form which is interpreted as being affected by an operator present in an utterance.

5. Copied properties of operators
5.1. The frontness operator

An operator functioning at the *phonological level* is the one assigning front or back quality to a syllable. This is the so-called palatal harmony in Turkic, a suprasegmental phenomenon of syllabic harmony (Johanson 1991). Syllables in Turkic are classified as either front or back. Both consonants and vowels can participate in the task of signalling the front or back status of a given syllable. Full vowels are either front or back and have, as a rule, a clear signal effect. Some consonants may also function as forceful back or front signals. For instance, while *k*, *g*, and *l* are front signals, *q*, *ġ*, *γ*, and *ł* are back signals in most Turkic languages. Their phonetic realisation differs in the individual languages, but the phonological contrast between front and back remains constant. Thus, in the Turkish words *kül* 'ash' versus *quł* 'slave', both the consonants and the vowels participate in signalling the quality of the syllable. Not all consonants have, however, the phonetic properties needed to function as a frontness or backness signal. See the Turkish examples (4).

(4) a. πFRONT   {kul}   →   [kül]   'ashes'
        πBACK   {kul}   →   [quł]   'slave'

     b. πFRONT   {bun}   →   [bün]   'base'
        πBACK   {bun}   →   [bun]   'distress'

The front versus back classification of syllables plays an important role both with respect to the internal structure of lexical items and in determining the phonological shape of suffixes. The tendency of harmony is operative within syllables and may also apply across syllables within a word form.

The operator assigning front versus back quality to syllables in Karaim has copied some phonological properties of Slavonic languages. Most of the consonants in Slavonic languages have both palatalized and non-palatalized variants. The palatalized ones occur often in the environment of front vowels.

The Karaim operator assigning the value 'front' or 'back' to a syllable applies the Slavonic principle of selection, i.e., consonants are palatalized in the environment of palatal vowels. Consequently, the Karaim operator assigning front quality to a syllable assigns palatal quality to the vowel and palatalized quality to all consonants.

(5) *Syllabic harmony in Lithuanian Karaim*

Front		Back	
k'el'd'im		kaldïm	
come:PAST.1.SG		stay:PAST:1.SG	
'I came'		'I stayed'	
yùv'g'à		orunɣa	
house:DAT		place:DAT	
'to the house', 'home'		'to the place'	
k'ȯp	m'ȧ	uprax	mə
many	Q	cloth	Q
'many?'		'cloth?'	

As a consequence, the consonant inventory of Karaim has become extended by the addition of the copied set of Slavonic palatalized consonants. Compare the Turkish consonant inventory in (6) with the Karaim one in (7). The Karaim inventory includes a palatalized variant of each consonant.

(6) *Inventory of Turkish consonants (in Turcological notation)*

	Labial	Alveo-dental	Prepalatal	Postpalatal	Velar	Glottal
Stop	p, b	t, d		k, g	q, ġ	
Fricative	f, v	s, z	š , ž			
Nasal	m	n				h
Affricate			č, ǰ			
Glide			y			
Liquid			l, ł, r			

(7) *Inventory of Karaim consonants (in Turcological notation)*

	Labial	Alveo-dental	Prepalatal	Postpalatal	Velar	Glottal
Stop	p, p'	t, t'		k, k'		
	b, b'	d, d'		g, g'		
Fricative	f, f'	s, s'	š , š'			x, x'
	v, v'	z, z'	ž, ž'			ɣ, ɣ'
Nasal	m, m'	n, n'			ŋ, ŋ'	
Affricate		c, c'	č, č'			
			dž, dž'			
Glide			y			
Liquid			l, l'			
			r, r'			

This copying process has affected properties of the operator assigning intrasyllabic frontness versus backness features and resulted in a new inventory of consonants in Karaim. The copying process has first affected the operator. This becomes clear when comparing the Russian and the Karaim consonant inventories. In Russian *š*, *c*, and *ž* are always non-palatal and *šč* and *č* are always palatal. The Karaim system, on the other hand, also has palatal and non-palatal variants of these consonants, since any consonant can be in the scope of the operator; note the following Karaim examples.

(8) FRONT {išči} → [iš'č'i] 'worker'
BACK {konšu} → [konšu] 'neighbor'.

5.2. The Q-operator

The interrogative Q-operator in Karaim is, like in other Turkic languages, the particle *mi,* which is always atonic and cliticized to the questioned constituent; note the following Turkish examples.

(9)   *Turkish*
   a. Siz   eve   gidiyor   **mu**sunuz?
        you   home:DAT   go:IYOR.PRES   Q:2.PL
        'Do you go home?'

   b. Siz   eve   **mi**   gidiyorsunuz?
        you   home:DAT Q   go:IYOR.PRES:2.PL
        'Do you go **home**?'

   c. Siz   **mi**   eve   gidiyorsunuz?
        you   Q   home:DAT go:IYOR.PRES:2.PL
        'Do **you** go home?'

Lithuanian Karaim has preserved this Turkic syntax of the interrogative particle, but, as a syntactic innovation, it has also copied the scope properties of the Russian interrogative particle *li.* The particle is cliticized to the first accented word of the Russian clause; see (10).

(10)   *Russian*
    ...ljubiš   li   ty   svoju   rabotu?
    like:PRES.2SG: Q   you   own:FEM.ACC job:FEM.ACC

    Da, ja   očen'   ljublju   svoju   rabotu.
    yes I   very   like:PRES.1SG own:FEM.ACC job:FEM.ACC
    '... whether you like your job? Yes, I like my job very much.'

The same scope properties may be observed in Lithuanian Karaim as well. The interrogative particle *mi* follows the first word of the clause although this word is not questioned. See the first example in (11), in which the particle is used in a main clause, and the second example, in which it is attached to the first word of the subordinated clause. Note that, whereas the Karaim particle can be used both in main clauses and in subordinated clauses, the Russian particle is usually used only in subordinated clauses. Thus, not the Russian syntax but the scope properties of the operator have been copied.

(11) *Karaim*

Iš'iyn'i	**m'ə**	s'ův'ȧs'?
job:POSS.2SG:ACC: Q		like:PRES:2SG

'Do you like your job?'

Astrï	iš'im'd'ȧn'	b'iyȧn'ȧm.
very	job:POSS.1SG:ABL	be fond of:PRES.1SG

'Yes, I am very fond of my job.'

B'il'm'im	m'en'	**-m'ə**	k'ib'it'k'ȧ	barïm.
know:NEG:1SG	I	Q	shop:DAT	go:AOR:1SG

'I don't know whether I will go to the shop.'

5.3. Discourse operators

The most frequently copied operators assign discourse and pragmatic functions. The particle *to* is a discourse operator copied from Slavonic. It is used in (12) and (13) to mark a vague consecutive relation.

(12) *Karaim*

Son	n'eč'ik	čïxsa,
end	when	emerge:COND

to	k'er'ȧk	alma	n'in'd'ikol'ek	savut.
PARTICLE	need	take:INF	some	pot

'After they have emerged, then you need to prepare the pot.'

(13)

Da	n'eč'ik	is's'i	da	ombun'd'i	n'eč'ik	kuyaš
and	when	warm	and	so	when	sun

to	bu	savutnu	k'er'ȧk	čïɣarma	azbar	usnu.
PARTICLE	this	pot:ACC	need	take out	yard	POSTP.on

'And when it is warm and sunny outside, then you have to put the pot out in the yard.'

The last example of contact-induced changes in the operator system of Karaim is a specific strategy applied to mark a constituent as pragmatically prominent. The hypothesis that operators are especially targets for copying is corroborated by the fact that this construction is widely spread in a huge linguistic area. It is found in Slavonic languages, in Turkic languages spoken in Slavonic surroundings, e.g. in Azerbaijanian dialects, in Gagauz in Moldavia and Bulgaria, in Balkan Turkic dialects. It is used also in Hungarian, in Modern Greek and even in Khalaj in Iran.

The construction under investigation is characterized by the non-initial position of the subjunctor, see (14).

(14) π PROM   SUBJUNCTOR

Let us first look at this construction in Russian. Russian temporal clauses are normally introduced by a subjunctor such as *kogda* 'when', but, as the following examples illustrate, one or several nominal constituents of the subordinated clause can also precede the subjunctor. Note that no extraction from the clause takes place; the nominal constituent *zima* 'winter' is syntactically and semantically still a constituent of the subordinated clause.

(15) *Russian*
  a. **А**   **зима**   **когда**   **придет**,   время   летит.
     but   winter   when   comes   time   flies
     'But when the winter comes, the time flies.'

  b. **А**   **зима**   **у нас**   **когда**   **придет**,   время   летит.
     but   winter   by us   when   comes   time   flies
     'But when the winter comes by us, the time flies.'

  c. **Вот**   **крем**   **когда**   **приготовишь**,   тогда   и   доделаем   торт.
     well   cream   when   you prepare   then   and   make ready   cake
     'When you have prepared the cream, we will make the cake ready.'

  d. **Мы**   доделаем   торт,   **крем**   **когда**   **приготовишь**.
     we   make ready   cake   cream   when   you prepare
     'We will make the cake ready when you have prepared the cream.'

The constituent preceding the subjunctor may be the subject or any other nominal constituent of the clause. Several constituents may stand in the pre-subjunctor position. The subordinated clause may introduce the sentence as in the first example. It can, however, also take the final position, in which case the constituent *krem* 'cream' can be interpreted as a contrastive focus.

Also a Karaim temporal subordinated clause is normally introduced by the subjunctor *nečik* 'when', but the non-initial position of the subjunctor occurs as well.

(16) *Karaim*
  Kiš   **n'eč'ik**   **k'el'ät'**,
  winter   when   come:PRS

  to   šayarda   t'ež   vaxt   ašat.
  PARTICLE   town:LOC   quick   time   pass:PRS
  'When the winter comes, time passes quickly in the town.'

Corresponding constructions are also used in other Turkic languages spoken in Slavonic linguistic surroundings, such as Gagauz and Macedonian Turkish.

(17) *Gagauz in Moldova* (Menz 1996)
  Ilkin   **kolxoza**   **ačan**   **geldik**
  first   kolkhoz:DAT   when   come:PST:1PL

  o   zaman   ödärdilär   pek   islāx.
  that   time   pay:R.PST:3PL   very   good
  'The first time when we entered the kolkhoz they paid very well.'

(18) *Macedonian Turkish* (Matras 1996)
  Šu   araba   kimindir,   **bura**   **ne**   **duruyor?**
  that   car   who:GEN:PRT   here   what   stand:PRS
  'Whose car is that which is parked here?'

Another genetically non-related language of the area is Hungarian. The construction under investigation here is frequently used in spoken varieties of Hungarian.

(19) *Hungarian*

a. **A   tél   amikor   jön,**
   the   winter   when   come:PRS

   a városban   gyorsan   múlnak   a   napok.
   the town:INESS   quickly   pass:PRS3   the day:PL
   'When the winter comes, the days pass quickly in the town.'

b. **A   városba   hogy   beértünk   megéheztünk.**
   the   town:ILL   that   arrive:PST:1PL become hungry:PST:1PL
   'When we arrived at the town, we became hungry.'

The internal processes contributing to the development of these constructions in the particular Slavonic, Turkic, Finno-Ugric languages, respectively, are surely different. A model construction or several model constructions have been successively copied into the languages of the area. The copies have been adjusted to the system of the particular languages or language varieties and are thus clearly never identical. The syntactic frames of the copying languages have been changed to the effect that these languages have become syntactically more similar, i.e. convergence has taken place.

6. Copying of operators and discourse strategies

I have presented here the hypothesis that operators are especially attractive for copying. This is not surprising considering that they pay a central role in the discourse strategies applied by speakers in speech production.

References

Csató, Éva Ágnes. (1994). On word order differences between Turkish and Karaim. *Dilbilim Araştırmaları*, 1994, 54-61.

Csató, Éva Ágnes. (1996). Some typological properties of North-Western Karaim in areal perspectives. Boretzky, Norbert, Werner Enninger and Thomas Stolz (eds) *Areale, Kontakte, Dialekte. Sprache und ihre Dynamik in mehrsprachigen Situationen*, Bochum-Essener Beiträge zur Sprachwandelforschung 24, Bochum: Brockmeyer, 68-83.

Csató, Éva Ágnes. (In print). Some typological features of the viewpoint aspect and tense system in spoken North-Western Karaim. Dahl, Östen (ed.) *Tense and aspect in the languages of Europe,* Berlin, New York: Mouton de Gruyter.

Csató, Éva Ágnes. (In print). Syntactic code-copying in Karaim. Dahl, Östen and Maria Koptjevskaja-Tamm (eds) *The Circum-Baltic languages: their typology and contacts.*

Csató, Éva Ágnes and Lars Johanson (1995). Zur Silbenharmonie des Nordwest-Karaimischen. *Acta Orientalia Academiae Scientarium Hung.* 48, 329-337.

Johanson, Lars. (1991). On syllabic frontness oppositions in Turkic. *Varia Eurasiatica,* Szeged, 77-94.

Johanson, Lars. (1992). *Strukturelle Faktoren in türkischen Sprachkontakten.* Stuttgart: Steiner.

Johanson, Lars. (1993). Code-copying in immigrant Turkish. Extra, Guus and Ludo Verhoeven (eds) *Immigrant languages in Europe,* Clevedon, Philadelphia, Adelaide: Multilingual Matters Ltd., 197-221.

Johanson, Lars. (1996). Kopierte Satzjunktoren im Türkischen. *Sprachtypologie und Universalienforschung* 49, 1-11.

Matras, Yaron. (1996). Layers of convergent syntax in Macedonian Turkish. Paper given at the workshop *Turkish in Contact, February 5-6, 1996.* Wassenaar: Netherlands Institute for Advanced Study in the Humanities and Social Sciences.

Menz, Astrid. (1999). *Gagausische Syntax. Eine Studie zum kontaktinduzierten Sprachwandel.* Dissertation. Universität Mainz.

Thomason, Sarah Grey and Terrence Kaufman (1988). *Language contact, creolization, and genetic linguistics,* Berkeley, Los Angeles, Oxford: University of California Press.

Affix-placement variation in Turkish

JEFF GOOD and ALAN C. L. YU
University of California, Berkeley

0. Introduction

According to the standard descriptions, complex Turkish nouns and verbs are formed by adjoining multiple suffixes, in a specific order, to a basic root. For example, the suffixes used to form the complex noun in (1) can only occur in that particular order.

(1) *ev*   *-ler*   *-in*   *-den*
    'house'   PL   3rd.POSS   ABL
    'from his house'

However, certain tense, mood, and aspect (TMA) markers regularly occur in different orders relative to each other.

(2) a. *bɪrak*   *-tɪ*   *-ysa*   *-m*
     'leave' PAST   COND 1st.SG
     'if I left ...'
    b. *bɪrak*   *-sa*   *-ydɪ*   *-m*
     'leave' COND PAST   1st.SG
     'if I would have left ...'

The fact that the two different suffix orderings in (2) produce verbs with different meanings shows that the ordering of the TMA suffixes above is not truly variable but is a function of morphology and semantics. However, it turns out that there are some limited cases where there is true variable suffix ordering in Turkish. Personal agreement endings, like *-m* in the verbs in (3), do not necessarily have to occur in their canonical position at the end of the verb (3a). It has been observed (Sezer 1998) that they can also be in penultimate position (3b) without producing any difference in meaning.

(3) a. *bɪrak*   *-tɪ*   *-ysa*   **-m**
     'leave' PAST   COND   **1st.SG**
     'if I left ...'
    b. *bɪrak*   *-tɪ*   **-m**   *-sa*
     'leave' PAST   **1st.SG** COND
     'if I left ...'

Variable ordering of the personal endings is not completely unconstrained however. The data in (4) shows that the combination of the progressive marker and the past marker only allows the personal ending in word-final positions.

(4) a. *oyn   -uyor -du   **-k***
    'play' PROG PAST **1ˢᵗ.PL**
    'we were playing'

  b. **oyn   -uyor -k   -tu*
    'play' PROG **1ˢᵗ.PL**   PAST

The purpose of this paper is to examine the restrictions on the variable ordering of personal endings in Turkish. We begin with an overview of Turkish verbal morphology in section 1. Personal agreement suffix placement variability and its formal analysis in terms of Optimality Theory (Prince & Smolensky 1993; McCarthy and Prince 1995) are presented in section 2. The distribution and the formal analysis of the Turkish plural marker *–lEr* is discussed in section 3. Theoretical implications of the analysis and concluding remarks are given in sections 4 and 5.

1. Turkish verbal morphology
1.1 Overview
The basic morphological structure of the Turkish verb is a verb root followed by a minimum of one, and possibly several, suffixing morphemes. These mark for tense, aspect, mood, and subject. In (5a), a basic schema is given with optional suffixes marked in parentheses. Underneath in (5b) is a verb that uses a variety of these markers.

(5) a.	**Root**	**(Negation)**	**Aspect**	**(Tense/Mood)**	**Personal Ending**
b.	*git*	*-m*	*-iyor*	*-du* *-ysa*	*-m*
	'go'	NEG	PROG	PAST COND	1ˢᵗ.SG
	'If I were going ...'				

The schema in (5a) shows the personal agreement ending as being verb-final because the personal ending always can, and for the most part does, appear in that position.

1.2 Personal endings
Turkish has four personal agreement paradigms. Here, we are only concerned with the two below which we have designated the *z*-paradigm and the *k*-paradigm in reference to the consonant in their first person plural forms.

(6)		**z-paradigm**		**k-paradigm**	
		Singular	**Plural**	**Singular**	**Plural**
	First	*-(y)Im*	*-(y)Iz*	*-m*	*-k*
	Second	*-sIn*	*-sInIz*	*-n*	*-nIz*
	Third	Ø	Ø	Ø	Ø

The *k*-paradigm endings only follow the two suffixes in (7a). The *z*-paradigm endings follow most other TMA suffixes. The TMA suffixes that take the *z*-paradigm that are referred to in this paper are given in (7b).

(7) a. Suffixes taking *k*-paradigm endings.
    -DI   Past (PAST)
    -sE   Conditional (COND)

   b. Suffixes taking *z*-paradigm endings.
    -Iyor   Progressive (PROG)
    -(y)EcEk Future (FUT)
    -Er   Aorist (AOR)
    -mIş   Unwitnessed/Evidential (EVID)

2. Personal ending ordering variability and its formalization
2.1 Data regarding personal ending ordering variability

Turkish grammars and textbooks (e.g. Sezer 1998, Underhill 1976) mention the possible ordering variability seen in (2) and (3). Sezer gives the following partial paradigm of the verb *görmek* 'to see' illustrating that the ordering variability can occur with the *k*-paradigm personal endings (8). Such variability is not possible for the *z*-paradigm suffixes.

(8) a. *gör-dü-yse-**m***   *gör-dü-**m**-se*   'if I saw . . .'
   b. *gör-dü-yse-**n***   *gör-dü-**n**-se*   'if you (sg.) saw . . .'
   c. *gör-dü-yse-**k***   *gör-dü-**k**-se*   'if we saw . . .'
   d. *gör-dü-yse-**niz***   *gör-dü-**nüz**-se*   'if you (pl.) saw . . .'
    'see'-PAST-COND-**PER**   'see'-PAST-**PER**-COND

As stated above, according to most authors (cf. Lewis 1967, Sezer 1998, Underhill 1976), there is no semantic difference between the pairs of verbs above. Kuroglu (1986), however, does claim that there is a slight difference of scope in the meanings of the above pairs. Our informants, nonetheless, agreed with most authors in believing the forms were equivalent in meaning, and we will be analyzing their dialects.

In addition to those patterns reported above, we found that Turkish also allows variable personal ending ordering in sequences such as the past perfect (*-DI* + *(y)DI*) (9a) and the past conditional (*-sE* + *-(y)DI*) (9b).

(9) a. Past Perfect
    *ara-dı-ydı-**m***   *ara-dı-**m**-dı*   'I had looked for'
    *ara-dı-ydı-**n***   *ara-dı-**n**-dı*   'you (sg.) had looked for
    *ara-dı-ydı-**k***   *ara-dı-**k**-tı*   'we had looked for'
    *ara-dı-ydı-**nız***   *ara-dı-**nız**-dı*   'you (pl.) had looked for'
    'look for'-PAST-COND-**PER**   'look for'-PAST-**PER**-COND

   b. Past Conditional
    *gel-se-ydi-**m***   *gel-se-**m**-di*   'I had looked for'
    *gel-se-ydi-**n***   *gel-se-**n**-di*   'you (sg.) had looked for'
    *gel-se-ydi-**k***   *gel-se-**k**-ti*   'we had looked for'
    *gel-se-ydi-**niz***   *gel-se-**niz**-di*   'you (pl.) had looked for'
    'come'-PAST-COND-**PER**   'come'-PAST-**PER**-COND

Grammatical Turkish verbs cannot be marked with more than two *k*-paradigm-taking verbal TMA markers. Given this restriction, there are four possible combinations of the two verbal suffixes which take the *k*-paradigm and, therefore, there are four potential cases to test for variable placement of *k*-paradigm personal endings. These are past+conditional (8), past+past (9a), conditional+past (9b), and conditional+conditional. This last combination is not allowed by Turkish semantics. The data above is, thus, fully representative of variable ordering of *k*-paradigm endings.

2.2.   A formal analysis

To capture the facts discussed so far formally, we propose to account for patterns such as (8) and (9) in terms of Optimality Theory.

The constraints employed here are defined in (10).

(10)   **MAX-M**(orph)   A morphosyntactic feature must have a realization in the output.
        LICENSE   A personal ending must be licensed by the immediately preceding suffix.
        ECON(omy)   Allow only one realization of a morphosyntactic feature. (Noyer 1993)

The LICENSE constraint is required to account for the facts described in (7) that some verbal TMA suffixes are always followed by *z*-paradigm endings while the rest are always followed by *k*-paradigm endings. For LICENSE to be effective, we must assume some sort of blocking or allomorph selection mechanism (cf. Riehemann 1993, Jurafsky and Koenig 1994, or Orgun 1996). A detailed description of such a mechanism, however, is beyond the scope of this paper (however, see Yu and Good [in press] for a morphosyntactic account of paradigm-selection). What is critical for this paper is simply that LICENSE will be violated by any forms in which the paradigm of the personal ending following a TMA suffix is not the one specified in (7).

The tableau in (11) illustrates how the constraints in (10) can account for the variable ordering described in (8) and (9). Because candidates in the tableau represent possible phonological forms, we use IPA notation for them. For ease of reference, however, outside of tableaus we have used the standard Turkish orthography.

(11)

giD, PAST, COND, 1st.PL	MAX-M	ECON	LICENSE
a. ☞ git-ti-jse-**k**			
b. ☞ git-ti-**k**-se			
c. git-**k**-ti-jse			*!
d. git-ti-jse-**jiz**			*!
e. git-ti-**k**-se-**k**		*!	
f. git-ti-jse	*!		

Forms (11a) and (11b) violate none of the proposed morphological constraints. Both are therefore acceptable outputs for the combination of the verb stem and morphological features given. Thus, we can account for why both forms can be

observed. A form like (11c) demonstrates how variable ordering is constrained. Verb stems do not license personal endings and, therefore, cannot be followed by them. So the suffix order in (11c) is not a grammatical output. Forms (11e) and (11f) are not possible simply because they violate one of the two constraints which, together, force outputs to realize all morphological features once and only once. Finally, a form like (11d) is ungrammatical since it has a z-paradigm ending following the suffix –(y)sE which licenses the k-paradigm.

In the tableau (12) we demonstrate how the constraints as defined account for the ordering variability of the past perfect as seen in (9a).

(12)

giD, PAST, PAST, 1st.PL	MAX-M	ECON	LICENSE
a. ☞ git-ti-jdi-**k**			
b. ☞ git-ti-**k**-ti			
c. git-**k**-ti-jdi			*!
d. git-ti-jdi-**jiz**			*!
e. git-ti-**k**-se-k		*!	
f. git-ti-jdi	*!		

Like in (10), the constraint rankings allow the two acceptable forms as outputs since they violate none of the morphological constraints. Other plausible candidates violate the constraints in (10) and are ungrammatical.

Our analysis of variable ordering cannot be considered complete yet. Only the k-paradigm endings permit variable ordering. When a verb contains two TMA suffixes which license the z-paradigm, the personal ending can only appear word-finally.

(13) a. *bul*   *-uyor* *-sun*
         'find'   PROG 2nd.SG
         'you are finding'
       b. *bul*   *-uyor* *-muş*   *-sun*
         'find'   PROG EVID 2nd.SG
         'you are apparently finding'
       b'. **bul*   *-uyor* *-sun*   *-muş*
         'find'   PROG 2nd.SG   EVID

The data in (13) reveals a deficiency in our constraints as defined so far. As can be seen in the tableau in (14), both (13b) and (13b') are predicted to be grammatical since neither violates LICENSE.

(14)

bul, PROG, EVID, 2nd.PL	MAX-M	ECON	LICENSE
a. ☞ bul-ujor-muʃ-**sun**			
b. ☛ bul-ujor-**sun**-muʃ			
c. bul-**sun**-ujor-muʃ			*!
d. bul-ujor-muʃ-**n**			*!
e. bul-ujor-muʃ	*!		

How can we account for the ungrammaticality of *z*-paradigm variable ordering and still maintain a system that allows us to analyze as grammatical the variable ordering of the *k*-paradigm? To answer this question, we need to take a closer look at the behavior of the two personal agreement paradigms. It turns out that their asymmetric behavior is not limited to variable ordering. In the next section, we show that the *z*-paradigm endings are actually clitics, while the *k*-paradigm endings are true inflectional suffixes.

2.3 A short digression: The clitic status of the *z*-paradigm
The claim that the *z*-paradigm personal endings are clitics is central to the rest of our analysis. We will present evidence from the phonological, morphological, and syntactic domains to support our argument. For ease of reference, the two paradigms are reproduced below (15).

(15)

	z-paradigm		**k-paradigm**	
	Singular	**Plural**	**Singular**	**Plural**
First	-(y)*Im*	-(y)*Iz*	-*m*	-*k*
Second	-*sIn*	-*sInIz*	-*n*	-*nIz*
Third	Ø	Ø	Ø	Ø

The most notable phonological difference between the two paradigms is that the *k*-paradigm endings can be stressed when they are word-final (16b) whereas *z*-paradigm endings can never be stressed (16a).

(16)   z-paradigm endings   k-paradigm endings
   a. *unut   acák   smɪz*   b. *unut   -tu   -núz*
      'forget'FUT   **2nd.PL**   'forget'PAST   **2nd.PL**
      'you (pl.)will forget'   'you (pl.) forgot'

Default Turkish stress is word-final. Thus, word-final *k*-paradigm markers behave as though they are truly suffixed to the word preceding them. One of the defining phonological features of clitics is an inability to be stressed (Sadock 1991, Zwicky 1985b). Thus, the phonological behavior of the paradigm endings is consistent with an analysis that the ones of the *z*-paradigm are clitics and those of the *k*-paradigm are suffixes.

In the morphological domain, *k*-paradigm endings behave like suffixes insofar as they can only attach to a very limited class of lexical items—namely, the past marker -(*y*)*DI* or the conditional -(*y*)*sE*. On the other hand, *z*-paradigm markers can follow not only the verbal markers in (7b) but also nouns and adjectives in

nominal sentences. This relative freedom of attachment is another characteristic of clitics (cf. Zwicky 1985a).

(17) a. *adam -sınız*   *adam -nız*
          'man' **2ⁿᵈ.PL**   'man' **2ⁿᵈ.PL**
          'you are men'
      b. *iyi   -yiz*   **iyi-k*
          'good' **1ˢᵗ.PL**   'good' **1ˢᵗ.PL**
          'we are good'

Finally, in the syntactic domain, there is evidence that z-paradigm endings behave as independent units in the syntax—another important property of clitics—whereas k-paradigm endings do not. There is a phenomenon in Turkish affecting conjoined verbs known as suspended affixation (cf. Lewis 1967, Orgun 1995) in which markers on the final verb in a conjunct have semantic scope over all the verbs in the predicate. For example, suspended affixation is observed in (18a) and the unsuspended counterpart is given in (18b). Note that the personal agreement suffixes in (18a,b) are from the z-paradigm. Personal agreement suffixes from the k-paradigm are prohibited from participating in this suspended affixation construction. In (18c) we have attempted to suspend a personal agreement suffix from the k-paradigm. However, this sentence is ungrammatical. The well-formed equivalent of (18c) is given in (18d) where both the conjoined verbs are marked for personal agreement.

(18) a. [*Hastane   -ye   gid -er,   onu   gör   -ür]   -üm*
          'hospital'   DAT   'go' AOR   3ʳᵈ.SG.ACC 'see' AOR   1ˢᵗ.SG

      b. [*Hastane   -ye   gid -er]   -im,   [onu   gör   -ür]   -üm*
          'hospital'   DAT   'go' AOR   1ˢᵗ.SG   3ʳᵈ.SG.ACC 'see'   AOR   1ˢᵗ.SG
          'I go to the hospital and see him.'

      c. *[*Hastane   -ye   git -ti,   onu   gör   -dü]   -m*
          'hospital'   DAT   'go' PAST   3ʳᵈ SG.ACC 'see'   PAST   1ˢᵗ.SG

      d. [*Hastane   -ye   git -ti]   -m,   [onu   gör   -dü]   -m*
          'hospital'   DAT   'go' PAST   1ˢᵗ.SG   3ʳᵈ.SG.ACC 'see'   PAST   1ˢᵗ.SG
          'I went to the hospital and saw him.'

Since the z-paradigm endings can take part in this syntactic reduction process, while the k-paradigm endings cannot, suspended affixation also suggests that the z-paradigm markers are clitics while the k-paradigm markers are suffixes.

2.4 Formalization of the behavior of the z-paradigm

Having established that the z-paradigm is a clitic paradigm, it is now possible to formalize its behavior by introducing a new constraint.

(19) **ALIGN-C**(litics)   Align(z-paradigm, R, word, R)
                            The z-paradigm must be the last morpheme of word.

The tableau in (20) illustrates how this new constraint accounts for the data in (13). Only (13b) is now ranked as grammatical and (13b') is excluded.

(20)

bul, PROG, EVID, 2nd.PL	MAX-M	ALIGN-C	LICENSE
a. ☞ bul-ujor-muʃ-**sun** (13b)			
b. bul-ujor-**sun**-muʃ (13b')		*!	
c. bul-**sun**-ujor-muʃ			*!
d. bul-ujor-muʃ-**n**			*!
e. bul-ujor-muʃ	*!		

The ranking of ALIGN-C above LICENSE is required to rule out ungrammatical forms like the one in (21) where a non-final suffix, the progressive marker *–Iyor*, licenses the *z*-paradigm but the final suffix, the past tense marker *–DI*, licenses only the *k*-paradigm.

(21) *gid*   -iyor   **-uz**   -du
 'go'   PROG 1st.PL PAST

The tableau in (22) shows how the ranking shown in (20) correctly rules out the forms in (21) and only allows the grammatical form *gidiyorduk* to surface.

(22)

giD, PROG, EVID, 1st.PL	ALIGN-C	LICENSE
a. ☞ gid-ijor-du-**k**		
b. gid-ijor-**uz**-du	*!	
c. gid-ijor-**k**-tu		*!

There is one last fact that needs to be captured in our analysis. The variable ordering we have seen in (8) and (11) clearly shows that the *k*-paradigm endings do not need to be word-final. However, there are cases where a non-final *k*-paradigm ending would be licensed but is ungrammatical. The data in (23) shows verbs which contain a conditional followed by an evidential suffix. The conditional *-(y)sE* licenses the *k*-paradigm while the evidential *-(y)mIş* licenses the *z*-paradigm.

(23) a. *yürü*   -se   -ymiş   **-sin**
   'walk' COND EVID 2nd.SG
    'if you are supposedly walking . . .'
   b. *yürü* -se   **-n**   -miş
    'walk' COND   2nd.SG EVID

The form in (23b) is ungrammatical even though it does not violate any sort of licensing constraint. The reason for this is an additional sort of licensing constraint which is sensitive to the final suffix of a verb. We define this constraint in (24)

(24) F(INAL) S(SUFFIX) A(GREEMENT): The agreement suffix on the verb must be from the agreement paradgim licensed by the last suffix of theverb.

Like LICENSE, FSA requires some sort of allomorph selection or blocking mechanism. The ranking of FSA is not crucial. For clearer exposition, we will give it the same ranking as LICENSE since they are both licensing constraints. (25) shows how our constraints account for the data in (23).

(25)

jyry, COND, EVID, 2nd.SG	ECON	ALIGN-C	LICENSE	FSA
a. ☞ jyry-se-jmiʃ-**sin**				
b. jyry-se-**n**-miʃ				*!
c. jyry-se-jmiʃ-**n**			*!	*
d. jyry-se-**sin**-miʃ		*!	*	
e. jyry-se-**n**-miʃ-sin	*!			

3. The plural marker *-lEr*

The constraints ranking proposed so far to account for the *k-* and the *z-*paradigms has the added advantage of accounting for another variable suffix ordering phenomenon in Turkish. The plural marker *-lEr* in Turkish is optionally used on verbs to mark for third-person plural subjects. (It is also optionally used on nouns to mark plurality.) Like the personal agreement endings, it shows possibilities for variable ordering on the verb (25a,b).

(25) a. *at   -acak -tɪ   -lar*
            'throw' FUT   PAST   **PL**
      b. *at   -acak -lar   -dɪ*
            'throw' FUT   **PL**   PAST
            'they were throwing'
      c. **at   -lar   -acak -tɪ*
            'throw' **PL**   FUT   PAST

Like all personal agreement endings, *-lEr* cannot follow verb roots (25c). Importantly, *-lEr* is not a member of any personal ending paradigm. It is simply a plural marker. It is clearly very different morphologically than the personal endings since it is optional and since it can attach to any noun (as opposed to the *z-*paradigm endings which can only attach to predicative nouns).

In the formalism developed here, it is not necessary to amend the analysis developed so far to account for the sort of ordering seen in (25). *-lEr* is simply taken to be a suffix licensed by all TMA markers. The tableau in (26) illustrates that the grammatical forms in (25a,b) are both judged grammatical using the constraint ranking developed above while other forms are not.

(26)

giD, PROG, PAST, **PL**	MAX-M	LICENSE
a. ☞ gid-ijor-du-**lar**		
b. ☞ gid-ijor-**lar**-dı		
c. git-**ler**-ijor-du		*!
d. gid-ijor-du	*!	

There is no grammatical verb form in Turkish which has three suffixes which license the *k*-paradigm. Thus, we could not test to ensure that variable ordering of personal endings would allow for a personal ending to be followed by more than one TMA suffix. However, there are verbs forms in Turkish with three TMA suffixes which license *-lEr*. It is the case with such verb forms that *-lEr* can appear after any of the three suffixes.

(27) a. *koş*   *-uyor* *-du*   *-ysa*   ***-lar***
      'run'   PROG PAST COND **PL**
   b. *koş*   *-uyor* *-du*   ***-lar***   *-sa*
      'run'   PROG PAST **PL**   COND
   c. *koş*   *-uyor* ***-lar***   *-dı*   *-ysa*
      'run'   PROG **PL**   PAST COND
      'if they were running . . .'

The data in (27) verifies our analysis by showing that, barring other constraints, verbal suffixes can follow any suffix which licenses them and variable ordering is not limited to the 'flip-flopping' of the last two suffixes seen in all the other examples. In (28), we show how our constraints predict the data in (27).

(28)

koʃ, PROG, PAST, COND, **PL**	MAX-M	LICENSE
a. ☞ gid-ijor-du-jsa-**lar**		
b. ☞ gid-ijor-dı-**lar**-sa		
c. ☞ gid-ijor-**lar**-dı-jsa		
d. git-**ler**-ijor-du-jsa		*!
e. gid-ijor-du-jsa	*!	

4. Theoretical implications
4.1 Level-ordering in the Turkish lexicon
Within the traditional theory of Lexical Phonology and Morphology (LPM: Kiparsky 1982), it has been argued that levels are strictly ordered with respect to each other. Inkelas and Orgun (1995, 1998), based on the interaction of the phonology and the morphology of Turkish, argue for a principle in LPM called Level Economy, according to which a form is subject to the phonology only of those levels at which it is morphologically derived—thus, they depart from the traditional, strictly serial approach to the lexicon.

The variable ordering of the plural suffix *-lEr* described here constitutes more evidence against traditional LPM. Based on the productive morphology and

phonology of Turkish, Inkelas and Orgun (1998) give the following organization to the Turkish lexicon:

(28)

Level:	1	2	3	4	5
Morphology	root	passive aspect relative negative	plural possessive	case	tense agreement interrogative

If LPM is applied in strict serial-order fashion, then it is predicted that the placement variation of *-lEr* is impossible. Pluralization, according to Inkelas and Orgun, takes place at level 3 while tense suffixation takes place in level 5. In order for (25ab) and (27) to be grammatical, strict level ordering must be relaxed and allow level 3 pluralization before *and* after tense suffixation. The pluralization evidence observed here concerns the interaction between morphological processes in the lexicon and is unrelated to Turkish morphophonology—thus this data offers support for Inkelas and Orgun's proposals from another area of Turkish grammar.

4.2 Morphological free variation

Cases of affixal placement variation have generally been analyzed as motivated by the phonotactics and the phonology of the language, such as the so-called *um-*infixation in Tagalog (McCarthy & Prince 1995) or mobile affixes in Huave and Afar (Noyer 1997). The suffix- placement variation discussed in this paper enriches the typology of affix-placement variation described since it is purely morphologically conditioned.

5. Conclusion

This paper illustrates the known variable affix ordering in Turkish: personal suffix placement and plural suffix placement. We argue that the different ordering behavior of the *k-* and the *z-*paradigms is a matter of lexical status. That is, the *z-*paradigm endings are argued to be clitics, while the *k*-paradigm endings are argued to be lexical suffixes. The phenomena presented here have repercussions for the strict level ordering approach to the theory of Lexical Phonology and Morphology.

References

Inkelas, Sharon and C. Orhan Orgun. (1995). Level ordering and economy in the lexical phonology of Turkish. *Language* 71:4.

Inkelas, Sharon and C. Orhan Orgun. (1998). Level (Non)ordering in recursive morphology: Evidence from Turkish. In S. J. Lapointe, D. K. Brentari & P. M. Farrell (eds.) *Morphology and its relation to phonology and syntax.* Stanford, CA: CSLI.

Koenig, Jean-Pierre and Dan Jurafsky. (1994). Type Underspecification and On-line Type Construction in the Lexicon. in Proceedings of WCFFL 13. CSLI: Stanford, CA, pp.270-285.

Kiparsky, Paul. (1982). Lexical morphology and phonology. In I.-S. Yang (ed.), *Linguistics in the Morning Calm.* Linguistics Society of Korea. Seoul: Hanshin. 2-91.

Kuruoglu, Güliz. (1986). Time reference in Turkish conditional sentences. In A. Akhu Koç and E. ErguvanlI Taylan (eds.), *Proceedings of the Turkish linguistics conference. August 9-10, 1984.* 129-43. Istanbul: Bogazici University.

Lewis, G.L. 1967. *Turkish grammar.* New York: Oxford.

Noyer, Rolf. (1993). Optimal words: Towards a declarative theory of word formation. Ms., Princeton University.

McCarthy, John & Alan Prince. (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.), *UMOP 18: Papers in Optimality Theory*, 249-384.

Orgun, C. Orhan. (1995). Flat vs. Branching structures in morphology: The case of suspended affixation. *Proceeding of the Berkeley Linguistics Society 21.*

Orgun, Orhan. (1996). Sign-based morphology: A declarative theory of phonology-morphology interleaving. Ph.D. dissertation, UC Berkeley.

Prince, Alan, and Paul Smolensky. (1993). *Optimality Theory: Constraint interaction in generative grammar.* Ms., Rutgers University and University of Colorado at Boulder.

Riehemann, Susanne. (1993). Word Formation in Lexical Type Hierarchies—A Case Study of bar-Adjectives in German. SfS-Report-02-93. Seminar für Sprachwissenschaft, Eberhard-Karls-Universität Tübingen. (Master's Thesis).

Sadock, Jerrold M. (1991). *Autolexical Syntax: A Theory of Parallel Grammatical Representations.* Chicago, IL: University of Chicago Press.

Sezer, Engen. (1994). *Yaşayan Türkçe.* Unpublished ms. Harvard University.

Underhill, Robert. (1976). *Turkish grammar.* Cambridge, MA: MIT Press.

Yu, Alan C. L. and Jeff Good. In press. Morphosyntax of two Turkish subject pronominal paradigms. To appear in NELS 30.

Tuvan Reduplication and Harmony

K. DAVID HARRISON
Yale University

0. Introduction

In this paper we present new data on reduplication in Tuvan, a Turkic language of south Siberia.[1] Output forms generated by this morphological rule reveal the operation of phonological constraints that are not apparent in the regular lexicon and morphology of Tuvan. We found that speakers exhibited considerable phonological variation in the production of reduplicants. We attribute this variability to speakers' indeterminate ranking of general markedness constraints in the language. Reduplication also yields output forms in which rounding harmony fails to apply even though all conditions for its application are present. We argue that certain of the constraints responsible for rounding harmony may emerge in novel contexts, though these constraints exert no observable effects anywhere else in the language. We claim that such constraints, which may remain dormant or inactive in the standard lexicon and morphology, can emerge and apply robustly in more marginal areas of the language such as reduplicants and loanwords. We shall refer to such constraints as being "inactive." Our investigation of Tuvan reduplication provides evidence for a rounding harmony constraint which is attested cross-linguistically, but which remains inactive in Tuvan and surfaces only under special conditions. We model the proposed constraints within the framework of Optimality Theory (Prince and Smolensky 1993) to show how variable rankings account for variation in output.

1. Tuvan reduplication

Tuvan has a morphological rule of full reduplication that signals either intentional vagueness or an informal, jocular register. The rule reduplicates the entire base while replacing the vowel of the initial syllable, usually with either [a] or [u] (hereafter referred to as "replacement vowels"). For reference, we present the Tuvan vowel inventory.

(1) Tuvan Vowels:

i	ü	ɨ	u		ii	üü	ɨɨ	uu
e	ö	a	o		ee	öö	aa	oo

Monosyllabic words are reduplicated as follows:

(2)

base	base + reduplicant	gloss
nom	*nom-nam*	'book'
er	*er-ar*	'male'
eet	*eet-aat*	'river delta'
is	*is-as*	'footprint'

ög	*ög-ag*	'yurt'
süt	*süt-sat*	'milk'
qɨs	*qɨs-qas*	'girl'
xol	*xol-xal*	'hand'
at	*at-ut*	'name'
aar	*aar-uur*	'heavy'

This simplified data set shows that all vowels except [a] are replaced by [a], while [a] is replaced by [u]. Except for the replacement vowel, the above reduplicants are identical to the bases from which they derive (we omit minor consonant alternations herein). We assume the base and reduplicant to be related by a faithfulness constraint (McCarthy and Prince 1995). This ensures that—modulo other constraints—they will resemble each other as closely as possible:

(3)   FAITH-BAS.RED   *the reduplicant must be identical to the base*

Apart from their replacement vowels, monosyllabic reduplicants are identical to bases. The reduplication rule must introduce a higher-ranked, anti-faithfulness constraint (cf. Kelepir 1999). This constraint, which must be specific to Tuvan, not universal, dictates that the first vowel of the base and replacement vowel must differ from one another:

(4)   *IDENT-BAS.RED(v1)   *replacement vowel must differ from base vowel*

These two constraints alone account for the data presented so far. But the reduplicant becomes considerably less faithful when the base is polysyllabic:

(5)	base	base + reduplicant	gloss
	idik	*idik-adɨk*	'boot(s)'
	inek	*inek-anak*	'cow'
	ulu	*ulu-alɨ*	'dragon'
	arɨ	*arɨ-uru*	'bee'
	nomdʒuur	*nomdʒuur-namdʒiir*	'read'
	oktaan	*oktaan-uktaan*	'throw'-PAST
	uduur	*uduur-adiir*	'sleep'-FUT

In these forms, vowels change in both initial and post-initial syllables. Changes to post-initial vowels of the base do not result directly from reduplication, but from the application of backness and rounding harmony to the reduplicant. Harmony constraints (described below) thus outrank base-reduplicant faithfulness. Reduplication 'feeds' backness and rounding harmony by providing a new initial-

syllable vowel which may potentially trigger harmony. For example, the reduplicant of the word for 'bee', *ari̇* , surfaces as *ari̇-uru*, not **ari̇-uri̇*. Reduplication may also 'bleed' Rounding harmony: in *ulu* 'dragon,' the rule introduces an unrounded replacement vowel and the reduplicant is thus *ulu-ali̇* rather than **ulu-alu*. In the absence of such feeding and bleeding effects, post-initial vowels do not change. The vowel [aa] in *oqtaan-uqtaan* 'threw'-REDUP remains unaltered because it is already [+back] and is not a potential target for rounding harmony (cf. section 3 below).

2. Backness harmony

Backness harmony operates on the natural vowel classes defined by the feature [+/-back], and requires all the vowels in a word to agree in backness. Since reduplicants always contain either [a] or [u] in the first syllable, it follows that they are always composed entirely of back vowels.[2] Backness harmony applies such that all front vowels in the base change to their back counterparts in the reduplicant. Thus *idiktiγ* 'boot' becomes *idiktiγ-adi̇qti̇γ*.

3. Rounding harmony

Tuvan rounding harmony (hereafter RH) requires a high vowel to be rounded if it follows a rounded vowel. RH makes reference to the features [+round] and [+/-high]. In an autosegmental analysis (Clements and Sezer 1982), the feature [round] may be construed as spreading rightward from any round vowel to any adjacent high vowel. RH targets high vowels [i] and [i̇] and yields high rounded vowels [ü] and [u] as outputs. RH obtains robustly without regard to height or backness of trigger vowels. Any rounded vowel serves as a trigger. Allophonic variations in high vowel suffixes such as the third singular possessive /-(z)V/ demonstrate most clearly the operation of rounding harmony (suffix vowels targeted by RH are underlined):

(6)

lexeme	suffix		gloss
atʃa	*-zi̇*	* *-zu*	'father'-3
iz	*-i*	* *-ü*	'footprint'-3
tool	*-u*	* *-i̇*	'story'-3
xol	*-u*	* *-i̇*	'hand'-3
xöl	*-ü*	* *-ü*	'lake'-3
bürü	*-zü*	* *-zi*	'wolf'-3

Were RH to under- or over-apply, outputs such as **iz-u* or **xol-i̇* would be possible. In fact, such forms are never attested in the lexicon and morphology of Tuvan and are judged by speakers to be ill-formed.

Tuvan RH exhibits the property of unboundedness (Anderson 1980). [+round] spreads any distance within a word span to adjacent target vowels (underlined):

(7)   *xögʒüm-tʃü-nü*   'music'-AGENT.ACC

       tool-dʒu-luγ   'story'-AGENT.COMIT

We stated earlier that RH obtains robustly without regard to height or backness of trigger vowels. Crucially for our analysis, RH also obtains without regard to length of target or trigger (targets underlined).

(8)

trigger	target	word	gloss
long	short	*ooʒu̲m*	'slow(ly)'
long	short	*boo-zu̲*	'gun'-3
short	short	*udu̲p*	'sleep'-CV
short	short	*ulu-zu̲*	'dragon'-3
short	long	*uru̲u̲*	'daughter'-3
short	long	*idii-udu̲u̲*	'boot'-3-REDUP-3
long	long	*tooru̲u̲*	'sunflower.seed'-3

These facts suggest that length plays no apparent conditioning role in the application of Tuvan RH. This is not the case in the Tungus languages Bayinna Oroch (Li 1996) and Evenki (Nedjalkov 1997), or in Daur Mongolian (Wu 1996) where long vowels fail to trigger RH, although they consistently undergo it. Shortness of trigger vowels thus seems to be a factor favoring harmony. We know of no languages where long vowels trigger RH but short vowels do not. The relationship between trigger goodness and length is implicational: if long vowels are good RH triggers in a given language then short vowels are too. The reverse is not true. A constraint that spreads rounding from short vowels would thus seem to be universally ranked above one that spreads from long vowels. We propose the following constraints to account for quantity-sensitive RH as attested in Evenki, Bayinna Orochen and Daur:

(9) Quantity-sensitive RH constraints
   i.   SPREAD[RD]/V   *Spread [+round] from a short vowel*
   ii. SPREAD[RD]   *Spread [+round]*

In the normal application of Tuvan RH, there is no evidence for the separability of these two constraints. They pattern together in all areas of the lexicon and affixal morphology, spreading rounding from all rounded trigger vowels. Since the set of constraints is thought to be universal, however, we should conclude that (9i) and (9ii) do indeed remain distinct in Tuvan, but never pattern distinctly. It is uniquely in the context of reduplication, as we will show, that these constraints may fail to apply in an identical manner.

4. Variability in output
The true complexity of Tuvan reduplication becomes apparent in the many dialect and individual versions of the rule. The author documented reduplication patterns for thirty-five individual speakers. Three types of data were collected: (i)

spontaneous instances of reduplication from speech; (ii) elicited reduplicated forms; (iii) speakers' judgments of well-formedness of reduplicants proposed by the author or produced by other speakers. Speakers' production of reduplicants may be classified into one of three basic patterns (dialect groups) according to the number of possible replacement vowels they employ (either two or three), and the degree of flexibility in mapping inputs to output. In the schema below, bold lines denote more robust mapping patterns, while plain and dotted lines denote less robust ones.

(10)   Mapping of base (input) vowel to replacement (output) vowel

	base vowel		replacement vowel
Dialect A	[a]	⟶	[u]
	[i ü e ö ɨ u o]	⟶	[a]
Dialect B	[a]	⟶	[u]
	[i e ö ɨ o]		
	[ü u]	⟶	[a]
Dialect C	[a]	⟶	[u]
	[e ö]		[o]
	[i o ɨ]		
	[ü u]	⟶	[a]
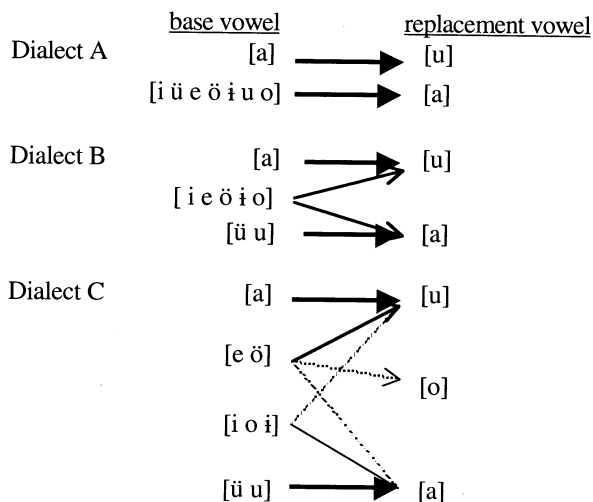
Dialect A (spoken in Kyzyl, Tuva), has only two replacement vowels and invariable mapping of inputs to outputs. Speakers of this dialect do not accept any alternative outputs as well-formed. Dialect B (spoken in central Tuva) also has two output vowels, but speakers show flexibility in mapping mid vowels and high unrounded vowels to either [a] or [u]. Speakers usually have an opinion about whether [a] or [u] is better, with an overall preference for [a]. Dialect C, (of the Süt-xöl region of Tuva), includes [o] as a replacement vowel, and shows even greater flexibility in mapping input to output. Dialect A is maximally simple and therefore less interesting, while dialect C has not yet been fully documented and is left for further research. The remainder of this paper will thus be primarily concerned with dialect B, which includes a number of regional variants of spoken Tuvan. For a speaker of dialect B or C, more than one well-formed reduplicant may be derived from a single base. Under (11) we see some of the various reduplicants produced by a single speaker of dialect B.

(11)

base	base+reduplicant I	base+reduplicant II	gloss
xol	*xol-xul*	*xol-xal*	'hand'
ök	*ök-ak*	*ök-uk*	'button'
is	*is-as*	*is-us*	'footprint'
er	*er-ar*	*er-ur*	'male'
öörenip	*öörenip-aaranip*	*öörenip-uuranip*	'study'-CV
ekkep	*ekkep-ukkap*	*ekkep-akkap*	'bring'-CV
eʒim	*eʒim-uʒum*	*eʒim-aʒim*	'friend'-1

The relation of input (base) vowels to output (replacement) vowels are shown in the histogram (12), which represents the composite output of ten speakers of dialect B (421 tokens). The horizontal axis shows the base vowels which serve as input to the rule. The vertical bars show outputs (replacement vowels) for each given input. Dark colored bars show the percentage of tokens for which [a] emerged as replacement vowel, and gray colored bars show how often [u] emerged.

(12)   Choice of replacement vowel for ten speakers (dialect B)
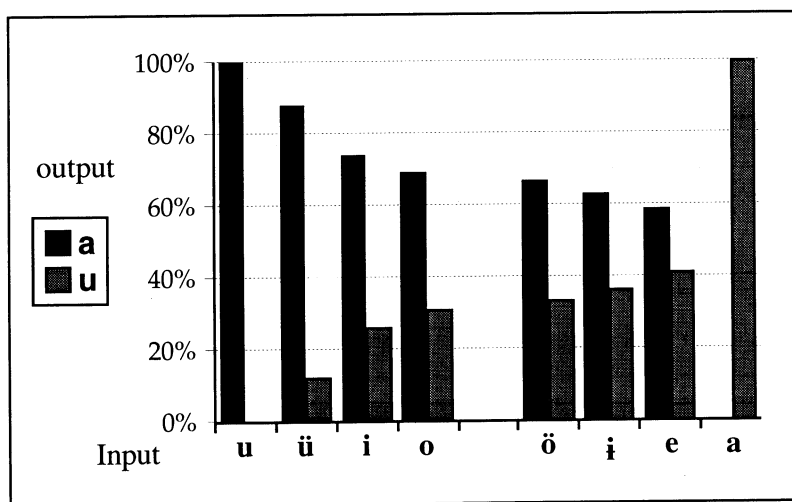
When the base contained [u], the reduplicant had [a] one hundred percent of the time. The same proportion held true for input [a]. All other input vowels varied in their output. Speakers showed an overall preference for [a], but [u] also appeared robustly as an alternative. I argue that the ratio of [a] to [u] in the output reflects speakers' uncertainty as to the relative markedness of these vowels. It also reflects the fact that the reduplication rule requires at minimum two replacement vowels, so both [a] and [u] are valid choices, even though [u] is more marked.

5. Variability and markedness

The reduplication rule as we have formulated it does not specify the quality of the replacement vowel. This can be shown, we believe, to fall out from general principles of the grammar. A markedness hierarchy among vowels might best account for the proportional occurrence of [a] and [u]. Cross-linguistically, [a] is the least marked vowel, and thus it is not surprising that [a] emerges most frequently as output. But [a] alone is insufficient to satisfy condition (i), since at least two replacement vowels are needed. The next best choice appears to be [u], if we limit the output to back vowels. Since [o] and [ɨ] are both highly marked, the proposed markedness hierarchy [a] > [u] > [o] > [ɨ] for back vowels is at least a plausible one.[3] The exact realization of markedness principles in Tuvan is left for further research. For now, we represent Tuvan markedness as a single constraint, *MARKED, requiring speakers to avoid marked segments.

6. The status of [round] in Tuvan

Front rounded vowels, of which Tuvan has two, are rare from a cross-linguistic perspective (Maddieson 1984). In a vowel inventory that has both front rounded vowels and back unrounded vowels, the feature [round] is not reinforced by the feature [back] as in English. Such reinforcement is referred to as enhancement (Stevens, Kaiser and Kawasaki 1986). Enhancement of one feature by another may serve to make contrasts among vowel qualities more salient. We propose that distinctive rounding without backness enhancement is a potentially difficult feature, not only in Tuvan but in languages in general (Kaun 1995).

Languages may employ various strategies for dealing with difficult contrasts. First, languages may employ positional neutralization (Steriade 1995) to limit such contrasts to prominent syllables. Tuvan, like most Turkic languages, restricts contrastive rounding to initial syllables. Low rounded vowels [ö] and [o] never appear post-initially. High rounded vowels [ü] and [u] appear post-initially only when generated by harmony, making rounding a fully predictable feature in post-initial syllables.

A second strategy for difficult features is to extend their domain by spreading them. This can be done with vowel harmony, which spreads a feature across the word span. It is probably not a coincidence that languages that spread [+round] via rounding harmony tend to have both front rounded vowels and back unrounded vowels in their inventories. The spreading strategy may thus be advantageous in vowel systems such as Tuvan where rounding is not enhanced by backness.

7. Variation in applying RH

We conclude that the constraints responsible for rounding spread are virtually undominated in Tuvan. Nowhere in the lexicon, borrowed lexicon, or morphology does RH fail to apply when the triggering conditions are met. Given the robust and pervasive nature of Tuvan RH, it is somewhat surprising that the output of reduplication provides unique examples of forms where rounding harmony fails to apply even though conditions for it appear to be present.

(13)

base	base+reduplicant I	base+reduplicant II	gloss
	RH obtains	**RH fails**	
azɨ	azɨ-uzu	* azɨ-uzɨ	'if'
arɨ	arɨ-uru	* arɨ-urɨ	'bee'-3
aazɨ	? aazɨ-uuzu	aazɨ-uuzɨ	'mouth'-3
aarɨ	? aarɨ-uuru	aarɨ-uurɨ	'burden'-3

For all speakers, *azɨ* 'if' is reduplicated as *azɨ-uzu*, with obligatory rounding harmony. But the nearly homophonous form *aazɨ* 'mouth'-3 is, for many speakers, reduplicated as *aazɨ-uuzɨ* without rounding harmony. Underapplication of RH can occur only if the triggering vowel is long. This effect is quite robust in reduplicants, but does not surface anywhere else in Tuvan.

Viewed from a cross-linguistic perspective, rounding harmony is frequently seen to be conditioned by various phonological features of the target and trigger (e.g. height or backness). The highly conditioned nature of many harmony systems can be best modeled as a family of constraints. These constraints may refer to various features of targets and triggers.

We suggested earlier that rounding harmony may be perceptually motivated as a strategy for dealing with a difficult contrast such as unenhanced rounding. We will now capitalize on that notion by proposing that it may be more crucial to spread the feature [round] from a short vowel than from a long one. A long vowel has greater perceptual salience, and may be better able to support a difficult contrast. This hypothesis provides the motivation for the two RH constraints proposed in (10).

Our formal model of Tuvan reduplication and RH relies on five constraints:

(14)   *IDENT-BAS.RED(V1)   *base vowel must differ from replacement vowel*
       FAITH-BAS.RED   *reduplicant should be identical to base*

       *MARKED   *avoid marked segments (*ɨ, *o)*
       SPREAD[RD]/V   *spread [+round] from a short vowel*
       SPREAD[RD]   *spread [+round]*

The full range of attested Tuvan reduplication patterns can be readily generated by the appropriate rankings of these constraints. For example, speakers showed three different patterns of RH from a long trigger:

(15)

		base	base+reduplicant	
speaker 1	(RH)	taaqpɨ	taaqpɨ- tuuqpu̱	'tobacco'
speaker 2	(no RH)	taaqpɨ	taaqpɨ-tuuqpɨ	
speaker 3	(optional RH)	taaqpɨ	taaqpɨ- tuuqpu̱ or taaqpɨ-tuuqpɨ	

Speaker 1 represented a minority of those we surveyed. His spreading of round was quite consistent regardless of the length of the trigger. We model this behavior with the following constraint ranking:

(16)   Speaker 1 **always spreads** rounding from a long trigger vowel in reduplicants. SPREAD[RD] outranks FAITH-BAS.RED, while SPREAD[RD]/v plays no decisive role:

aazɨ 'mouth'-3	*ɨ, *o	*IDENT-BAS.RED (V1)	SPREAD [RD]/V	SPREAD [RD]	FAITH-BAS.RED
aazɨ-aazɨ		* !			
aazɨ-uuzɨ				* !	*
aazɨ-uuzu					* *
aazɨ-aaza		* !			*
aazɨ-ɨɨzɨ	* !				* *

Most speakers we surveyed were like speaker 2, in that they always applied RH when the trigger was short, but consistently failed to apply RH when the trigger was long. The effect we describe is thus quite robust, though absent from the regular lexicon and morphology of the language.

(17)   Speaker 2 **never spreads** rounding from a long trigger vowel in reduplicants. FAITH-BAS.RED outranks is interspersed between the RH constraints.

aazɨ 'mouth'-3	*ɨ, *o	*IDENT-BAS.RED (V1)	SPREAD [RD]/V	FAITH-BAS.RED	SPREAD [RD]
aazɨ-aazɨ		* !			
aazɨ-uuzɨ				*	*
aazɨ-uuzu				* * !	
aazɨ-aaza		* !		* *	
aazɨ-ɨɨzɨ	* !			* *	

Finally, speaker 3 **optionally spreads** rounding from a long trigger vowel in reduplicants. We found such speakers consistently able to alternate between forms with RH and forms without RH, thus employing both of the above rankings.

In the case of reduplicants with short vowel harmony triggers, all speakers consistently applied RH, and judged proposed forms without RH to be ill-formed.

(18)   Speakers **always** spread rounding from short trigger vowels in reduplicants. SPREAD[RD]/V outranks base-reduplicant faithfulness, while SPREAD [RD] plays no active role.

azɨ 'if'	* ɨ, *o	*IDENT-BAS.RED (v1)	SPREAD [RD]/V	SPREAD [RD]	FAITH-BAS.RED
azɨ-azɨ		* !			
azɨ-uzɨ			* !		*
azɨ-**uzu**					* *
azɨ-aza		* !			*
azɨ-ozɨ	* !		*		*
azɨ-ozu	* !				* *
azɨ-izɨ	* !				*

8. Discussion

Tuvan speakers have no evidence from the regular lexicon and morphology of their language that RH may fail to obtain when the potential trigger is long. Such cases are unattested in the standard language. Yet in novel contexts, when called upon to apply the reduplication rule, speakers produce forms in which RH fails to apply precisely when the trigger is long. The underapplication of RH when all conditions for it are present is clearly an anomaly in Tuvan. This output reflects "on-line" decisions made by speakers about the ranking of RH constraints relative to faithfulness constraints which are otherwise inactive.

We propose that phonotactic constraints that may be entirely dormant or have no active role in the native lexicon or morphology of a language can nonetheless surface and play an active role in more marginal or novel areas of a language such as word games, reduplicants, and loanwords. When these hidden phonological constraints do come into play, speakers may exhibit some uncertainty about their ranking, or may assign them a novel ranking. The assigned ranking can have the effect of demoting other constraint(s) that would otherwise be dominant.

In the case discussed herein, a sizeable number of speakers appears to be able to demote one of the RH constraints, causing RH to fail in some cases. In demoting this constraint, they effectively break up the family of RH constraints, which everywhere else in the language pattern as a uniform block. This raises theoretical issues about the behavior and (non)uniform patterning of constraint families.

The generally dormant or inactive status of the base-reduplicant faithfulness constraint and the fact that it becomes active only under reduplication, may account for speakers' uncertainty about its ranking. Clearly, no consensus has yet been reached in the speech community about where this constraint should rank relative to RH constraints. Dialectal and individual differences in ranking persist. Furthermore, many speakers seem content to employ at least two alternative rankings on different occasions, and do not seem to be in rush to assign a ranking. This raises important issues about the stability of variable constraint rankings and speakers' tolerance for uncertainty.

Notes

[1] Support for fieldwork in Tuva in 1998 was provided by the International Research and Exchanges Board (IREX). Tuvan language data were provided by the following people, to whom I am deeply grateful: Ch. Dolgar-ool, K. Myldyk, A. Irgit, Ch. Kuular, A. Sarygbai, V. Süzükei, S. Ondar, M. Kongar-Sürüng, I. Dongak, U. Mongush and others.

[2] Exceptions to this generalization are disharmonic loanwords such as [radiyo] 'radio', which are common in Tuvan (Anderson and Harrison, forthcoming). The patterning of disharmonic forms under reduplication raises a number of important theoretical issues, which are discussed in a separate paper. See (Harrison, forthcoming)

[3] Establishing a markedness hierarchy for the vowels of a language is a difficult theoretical and empirical problem. In proposing a tentative hierarchy for Tuvan vowels, we considered the following factors: (i) frequency of segments across languages, (ii) frequency of segments in Tuvan, (iii) alternations in productive morphology, (iv) positional neutralization, (v) acoustic values (central vs. peripheral), and (vi) the facts of reduplication.

References

Anderson, Gregory D.S. and K. David Harrison. (forthcoming). *Tyvan*. Languages of the World/Materials, vol. 222. München: LINCOM EUROPA.

Anderson, Stephen R. (1980). Problems and perspectives in the description of vowel harmony. In R. Vago (ed.) *Issues in vowel harmony*, pp. 1-48. Amsterdam: John Benjamins.

Clements, George N. and Engin Sezer. (1982). Vowel and consonant disharmony in Turkish. In Harry van der Hulst and Norval Smith (eds.) *Structure of phonological representations* (Vol. II), pp. 213-255. Dordrecht, Holland: Foris Publications.

Harrison, K. David. (forthcoming). Harmony and underspecification in Tuvan. In *MIT working papers in Linguistics*.

Kaun, Abigail R. (1995). *The typology of rounding harmony: An optimality theoretic approach*. Ph.D. dissertation, UCLA.

Kelepir, M. (1999). Emphatic non-identical reduplication in Turkish. Handout, LSA Annual Meeting.

Li, Bing. (1996). *Tungusic Vowel Harmony*. The Hague: Holland Acad. Graphics.

Maddieson, Ian. (1984). *Patterns of sounds*. Cambridge UK: Cambridge Univ. Press.

McCarthy, John and Alan Prince. (1995). Faithfulness and reduplicative identity. In *University of Massachusetts occasional papers in linguistics 18: Papers in Optimality Theory*, pp. 249-384. Amherst, MA: GLSA.

Nedjalkov, Igor. (1997). *Evenki.* New York: Routledge.

Prince, Alan, and Paul Smolensky. (1993). Optimality theory: Constraint interaction in generative gramma*r. Technical Report #2 of the Rutgers Center for Cognitive Science*, Rutgers University.

Steriade, Donca. (1995). Underspecification and markedness. In John A. Goldsmith (ed.), *The handbook of phonological theory.* Cambridge, MA: Blackwell.

Stevens, Kenneth, S. J. Keyser and H. Kawasaki. (1986). Toward a phonetic and phonological theory of redundant features. In J. Perkell and D. Klatt (eds.), *Invariance and variability in speech processes,* pp. 426-449. Hillsdale, NJ: Lawrence Erlbaum.

Wu, Chaolu. (1996). *Daur.* Languages of the World/Materials vol. 93. München: LINCOM EUROPA.

Attractiveness and relatedness: Notes on Turkic language contacts

LARS JOHANSON
University of Mainz, Germany

0. Introduction

The aim of the present paper is to give a short account of some thoughts about possible ways of argumentation concerning the external relations of the Turkic languages, in particular the putative language family commonly referred to as 'Altaic'.

For the sake of brevity and clarity, I will dispense here with references to the vast literature on the subject and just refer to some previous work in which I have argued for the theoretical claims in more detail and on the basis of fuller empirical data.

1. Relatedness versus copying

Comparative linguistics disposes of a strong apparatus for judgments concerning cognate items on the basis of regular phonetic correspondences. However, the theoretical foundations for determining items as copied—or, with a traditional but inadequate term, "borrowed"—from other languages have so far been relatively weak. It will be argued here that stricter criteria are also required for the assessment of contact-induced similarities between languages.

The situation may be illustrated with the external relations of the Turkic languages (Johanson and Csató 1998), which exhibit indubitable similarities with Mongolic, Tungusic, Korean, and even Japanese. Historical linguists who do not believe in chance, i.e. accidental similarity, are confronted with two options in this case: to ascribe the parallels to *genetic relatedness*, or to *copying as a result of contact*.

2. The Altaic hypothesis

As regards the first option, the so-called "Altaic hypothesis" is still a matter of controversy. The existence of an Altaic protolanguage has not yet been proved unequivocally with the methods of comparative linguistics. This means that the Altaic languages have not been shown to be interrelated in the same sense as, for instance, Indo-European ones. This is even true of the relationship between Turkic and Mongolic. The difficulties are due to the enormous time dimensions involved in the reconstruction of the protolanguage in question and the poor documentary resources for a deep perspective of this kind, namely lack of older data, textual evidence, and significantly older written records of the languages concerned (Johanson 1990).

The second option means that the similarities are due to copying. It is beyond question that some of the languages involved, whether genetically related or not, have influenced each other for many centuries. Irrespective of their origin, immediate neighbors frequently share traces of areal interaction. It is often highly difficult to distinguish areal convergence from genetically conditioned divergence (Johanson, in print).

The advocates of the Altaic hypothesis, the so-called Pro-Altaicists, claim that the parallels are established on a set of regular phonetic and semantic

correspondences. They take the genetic relationship to be proven if connections of this kind can be demonstrated for a significant portion of the basic vocabulary.

The opponents of the Altaic hypothesis, the so-called Anti-Altaicists, explain the parallels as the result of copying. Pro-Altaicists answer that, in order to disprove genetic relatedness, it would indeed be necessary to demonstrate that the parallels are random and that there are no regular phonetic and semantic correspondences between the given languages. Anti-Altaicists maintain that it is also possible to establish regular correspondences between unrelated languages in cases of massive borrowing. Pro-Altaicists respond that massive copying of basic vocabulary is implausible.

The disagreement is understandable. Genetic relatedness cannot be observed directly, just the traces it may have left in the data available to the linguist. Different positions concerning the interpretation of such data are natural. For some reason, this is nevertheless a highly emotional domain (Johanson 1999c). Pro- and Anti-Altaicists are permanently at daggers drawn with each other. The discussion is full of sharp polemics, use of military terms, strict assignment to hostile camps, emotional attacks, hints at psychological roots of the controversy, etc. Being an Anti- and a Pro-Altaicist is equally blamable in the opposite groups. And agnostic positions are condemned in both camps. In spite of all the dangers in this minefield, a few comments on the genetic question will be ventured here.

Opponents of the Altaic hypothesis frequently urge its advocates to apply more stringent criteria. It would, however, be equally appropriate to direct this request to scholars who deal with copying in a rather loose way (Johanson 1995b). It is often observed that similar items that cannot be proven as cognates are declared "borrowings" without reference to any criteria whatsoever. Stringency is also needed in hypothesizing about copying processes. In order to find arguments for genetic relationship and contact, evidence for cognates and copies, it is advisable to utilize data from earlier research in historical linguistics and to pay attention to the findings of studies in language acquisition and psycholinguistics. Assumptions about correspondences due to copying should be tested in terms of empirical plausibility and diachronic naturalness.

3. Code-Copying

In the following, reference will be made to the so-called Code-Copying Model, an integrated descriptive framework for typological analysis of various contact-induced phenomena conceived of as linguistic 'copies' (Johanson 1992, 1993a, 1998, 1999b). The term *code-copying* is employed for the common interaction of linguistic codes mostly referred to as "borrowing". The crucial idea is that copies of elements from a foreign *model code* are inserted into a *basic code*, which provides the *morphosyntactic frame* for the insertion.

The choice of the term *copying* instead of *borrowing* is motivated by the important distinction between originals and copies. Copies are always part of the copying system and to some degree adapted to it; they are thus never identical with their originals. Code-copying always affects the structural characteristics of the basic code.

There are two kinds of insertion depending on the assignment of the two codes: *adoption* and *imposition*. In the case of adoption, speakers 'take over' copies from a secondary language into their primary language. In the case of imposition, speakers 'carry over' copies from their primary language into their

variety of a secondary language. The effects of adoption and imposition may be rather different.

In the case of *global copying*, a foreign item with all its properties (meaning, shape, combinability, or frequency) is the object of the copying. In the case of *selective copying*, only individual aspects of a foreign model are copied: meaning, shape, combinability, frequency (see, e.g., Johanson 1993b). Only global copying yields form-and-function items suitable as objects of comparison in genetic studies, i.e. as evidence against cognateness; e.g. the Tatar lexeme *qara-* 'to look' and the Tuvan verbal noun suffix *-(I)lda* are both globally copied from Mongolic.

4. Turkic language contacts

The Code-Copying Model has been used to describe typological tendencies in language contacts, e.g. how Turkic languages have influenced others and vice versa. This is a vast research area, since Turkic has been in close contact relations with numerous languages throughout its history. Strong mutual influences with Indo-European have arisen through a long-lasting Turkic-Iranian symbiosis in Central Asia. Other contact partners include Mongolic, Slavic, and Finno-Ugric varieties. The code-copying framework may serve to describe both ongoing contact phenomena such as the development of Turkish in Germany and processes that can only be observed in a diachronic perspective, e.g. Slavic influence on the Turkic Karaim language during its history of at least 600 years (see Csató, this volume).

There is abundant material to study the results of all these contacts. Some of them have been presented in a monograph on Turkic language contacts (Johanson 1992; English version forthcoming). The book tries to sum up what kinds of contact-induced changes have affected the structures of Turkic languages and their non-Turkic counterparts, what structural factors have been decisive for copying, what kinds of non-lexical items and properties have been copied under various social circumstances, etc.

5. Changes in the morphosyntactic frame

The study in question suggests that almost any feature can be copied given the appropriate social conditions. The morphosyntactic frame provided by the basic code may itself undergo essential changes through copying of new morphosyntactic features, e.g. markers of grammatical functions. Under conditions of sustained intensive contact, speakers may progressively restructure their basic code on the model of a dominant code. Successive frame changes pave the way for the insertion of further grammatical copies (Johanson 1999a).

This is sufficiently documented in the history of Turkic. The Turkic group includes several high-copying languages that display a good deal of frame in-novation due to copying from genetically unrelated and typologically different languages. Some go very far in altering their basic morphosyntactic frame. Due to long and intensive contact, Turkic varieties spoken in Central Asia and Iran have developed considerable similarities with Persian. Gagauz, predominantly spoken in southern Moldova and neighboring parts of Ukraine, has developed under strong Slavic influence. Karaim, now spoken in Lithuania and Ukraine, has also converged with Slavic contact languages. Salar, spoken in China, has copied numerous frame-changing elements due to sustained contact with Chinese and Tibetan. The same is true of some languages that have been heavily influenced by Turkic.

Note that such changes in the basic code have taken place without a changeover to the model code. Heavy copying may make the basic code very similar to the model code without leading to a code shift. The influenced languages have not been transformed into the languages that have influenced them. Thus, Ottoman Turkish remained Turkic irrespective of its overwhelming load of copies from Arabic and Persian. It copied practically everything except a few basic function markers. Karaim has remained Turkic in spite of all its frame-changing developments. High-copying codes remain identifiable by the non-copied core elements of the frame. Even if the consecutive changes a given language has undergone may have led to considerable deviations from the features typical of its genetic group, it may still be classified with that group.

6. Attractiveness

Evidence from contact linguistics may allow hypotheses about what parts of grammar are mostly affected in copying processes. The study mentioned above (Johanson 1992) discusses what makes structures relatively 'attractive', more susceptible to copying, more readily copiable than other features.

Most Turkic languages today exhibit attractive features in the sense of transparency and regularity. Throughout their history they have, at least in the central areas, tended to abolish and change "marked" structures, to promote salient semantic and material structures, to regularize paradigms, to substitute attractive features for unattractive ones. The dominant Turkic type of today is a result of continued reinforcement of regular and transparent structures. Even the earliest Turkic known to us—documented in 8th century inscriptions—may be the result of a leveling koinéization.

The study of well-known contact-linguistic cases suggests that one type of circumstantial evidence in the discussion of the Altaic hypothesis may be based on the concept of attractiveness. It might be used as an argument to strengthen or weaken the probability of an element having spread as the result of contact. Evidence for relatedness might be premised on features that have proved unattractive for copying. Empirical knowledge of what items are less likely to be copied may lead to hypotheses about prehistorical copiability.

7. The least copiable non-lexemic items

One crucial issue concerns the copiabililty of free and bound non-lexemic items. In order to gain insights of relevance for the questions of relatedness it should be determined empirically what morphological elements are most resistant to contact-induced frame changes. What elements refuse being replaced by copies in the frame-providing code? What native items are finally left intact even in high-copying languages? What items permit us to classify a language with a particular genetic group even if it strongly deviates from the features typical of that group?

Proponents of the Altaic relationship have identified a fair number of common morphological elements as putative cognates. But they have also characterized numerous unintelligible elements of word forms as 'suffixes', without accounting for their functions. Anti-Altaicists, on the other hand, almost never acknowledge any constraints on copying.

Copiable items of this kind often exhibit a salient semantic and material structure. Morphological opacity and irregularity are unattractive properties. The items copied have a relatively specific meaning and frequently a relatively elaborated shape. This may mean that they are mainly copied at early stages of

grammaticalization. Thus, relators such as postpositions are easily copied while they are still comparatively salient. The ones based on nouns belong to the most copiable relators; e.g. Turkish *sebebiyle* 'because of' is based on Arabic *sabab* 'reason' (see Johanson 1993c, 1996a). A further case in point is the high copiability of discourse-relevant conjunctions (Johanson 1997, 1998). Items that have arrived at the end of their grammaticalization path—exhibiting less salient structures, more reduced shapes, and more general, abstract meanings—seem to be less suitable candidates for global copying.

If we find corresponding items of less copiable kinds, i.e. similarities between Altaic languages in domains that are empirically less susceptible to influence, this might be an argument for genetic relatedness. Items that can be shown to be old and subject to long grammaticalization processes, provide particularly good arguments.

It may, however, be very difficult to distinguish between mere lookalikes and true cognates in this domain. The situation is never so simple that two putative cognates "look the same" and "mean the same". In the development of suffixes, otherwise valid phonological laws are often violated, and exceptional developments may be expected. Irregularities in bound morphemes are often removed through analogy, which may render reconstruction and proof of cognateness impossible. Items etymologically connected with each other may be at different stages of grammaticalization. High-copying languages possess several historical layers of copies reflecting different contacts. Each layer makes it increasingly difficult to distinguish between non-copied items and nativized copies.

8. Candidates for cognateness

Certain categories have proved less copiable in the documented history of Turkic.

(i) Relators such as case suffixes, in particular markers of central syntactic functions as the accusative, dative and genitive suffixes, are all of high age and display highly generalized meanings as well as less salient shapes. There are no known examples of case suffixes copied into Turkic varieties. However, as a result of imposition (see section 3) due to very long and intensive contact, Turkic case markers have been globally copied into North Tajik dialects.

(ii) Bound aspect-mood-tense markers also have abstract, generalized meanings. In the course of their functional development, they have often been replaced by other native items which have renewed the expression of their previous functions. However, there are no known examples of items globally copied from non-Turkic models.

(iii) Personal pronouns—as well as predicative suffixes going back to them—have been rather stable in most Turkic languages. There are no known examples of copying.

(iv) Markers of actionality and voice filling the position next to the primary verb stem have not been replaced by copies from other codes. Some of them have, however, vanished, and some have been replaced by periphrastic postverb constructions.

There are also other examples of native elements that have remained stable in the history of Turkic. Thus, copular verbs ('to be') and ancillar verbs (e.g. 'to do') have been replaced by new native verbs (e.g. *er-* > *tur-* 'to be', *qïl-* > *eyle-* > *et-* > *yap-* 'to do'), but not by copied ones.

Looking at the categories i-iv in a comparative Turkic-Mongolic perspective, we may observe the following:

(i) Certain case suffixes show similarities, e.g. locatives such as Mongolic *-dA* ~ vs. Turkic *-DA*, the Mongolic terminative in *-čA* vs. the Turkic equative in *-čA*. The explanation of the Ordos, Khalkha, and Kalmyk accusative suffixes (type *-IG*) as global copies of the Orkhon Turkic accusative suffix in *-(°)G* seems little convincing.

(ii) Similarities are also observed between certain aspect-mood-tense markers, which all have reduced shapes and must be of considerable age, e.g. the Mongolic habitual in *-dAg* vs. the Turkic participle in *-DOQ*. It is even possible that a finite predicative marker such as Mongolic *-ba* (*irebe* 'has come') may correspond to a converb marker such as Turkic *-b* (e.g. *kelib* 'having come') (Johanson 1995a, 1996b).

(iii) Pronouns display similarities, e.g. Mongolic *bi* vs. Turkic *ben* 'I'. The oblique stems contain a so-called pronominal *n*, e.g. Mong. *minu* < *binu* 'my'.

(iv) There are similarities between the old actionality and voice markers that fill the position next to the primary verb stem, e.g. voice markers such as causative suffixes. The Mongolic cooperative suffix *-lčA* may be connected with Turkic *-(°)š*.

9. Irregularities

Interestingly enough, several of the cases just mentioned are exactly the ones in which Turkic has preserved traces of older irregularities. The predominantly regular and transparent morphological structures of contemporary Turkic languages evoke the deceptive impression that they have no secrets to reveal. On closer inspection, we find cases of morphological opacity and irregularity that appear typologically inconsistent with the rest of the structure. Anyone familiar with modern Turkish will know phonologically unpredictable allomorphs of personal pronouns, of causative suffixes, and of the old present tense marker in *-r* ("aorist"). Personal pronouns display vowel mutation (Turkish *ben* 'I', *bana* 'to me', Early Turkic *ben* 'I', *bini* 'me'), a phenomenon at variance with the agglutinative language type.

These are less attractive features of older stages of development that have not been subject to regularization. Turkic has preserved them exactly in cases where we, on independent grounds, expect the lowest susceptibility to copying, and where we also find the most striking Altaic parallels.

10. Conclusion

Candidates for Altaic cognateness might well be searched for among the categories just mentioned. The similarities are too striking to be random; the parallels can hardly be accidental. While they do not offer conclusive proof of the Altaic hypothesis, they are indeed indicative of close and systematic relations. It is true that the functional parallels are not always quite clear and that there are also various phonetic problems which make it difficult to prove cognateness. On the other hand, it is difficult to think of these abstract and reduced elements as copied from foreign codes, unless the copying occurred at very early stages of grammaticalization. In any case, there is no reason to accept the uncritical assumption that grammatical items of this type are freely copied back-and-forth among unrelated languages.

Note that areal convergence is not, as has often been suggested in the literature, a valid argument against common genetic origin. Let us assume that Turkic and Mongolic are interrelated and that the earliest Turkic known to us (8th century) is a *koiné* with simplified, "attractive" structures. If this is the case, the earliest Mongolic known to us (13th century) can be expected to have preserved an older, more complicated stage of development. Later areal interaction has demonstrably led to considerable simplification in Mongolic. For example, the gender distinctions and the inclusive vs. exclusive opposition in the 1st person plural have more or less vanished.

Such cases of convergence between Turkic and Mongolic certainly do not prove conclusively that the language groups in question go back to a common ancestor that exhibited gender distinctions and an inclusive vs. exclusive opposition. It should, however, be equally clear that they do not disprove this possibility.

References

Csató, Éva Ágnes. (1999). Analyzing contact-induced phenomena in Karaim. *Proceedings of the special session of the 25th annual meeting of the Berkeley Linguistics Society* .Berkeley: Berkeley Linguistics Society.

Johanson, Lars. (1990). Zu den Grundfragen einer kritischen Altaistik. *Wiener Zeitschrift für die Kunde des Morgenlandes* 8, 103-124.

Johanson, Lars. (1992). *Strukturelle Faktoren in türkischen Sprachkontakten.* (Sitzungsberichte der Wissenschaftlichen Gesellschaft an der J. W. Goethe-Universität Frankfurt am Main, 29:5.) Stuttgart: Steiner. [English translation in preparation; London: Curzon.]

Johanson, Lars. (1993a). Code-copying in immigrant Turkish. Extra, Guus and Ludo Verhoeven (eds.) *Immigrant languages in Europe.* Clevedon, Philadelphia, Adelaide: Multilingual Matters, 197-221.

Johanson, Lars. (1993b). Zur Anpassung von Lehnelementen im Türkischen. Laut, Jens Peter and Klaus Röhrborn (eds.) *Sprach- und Kulturkontakte der türkischen Völker.* (Veröffentlichungen der Societas Uralo-Altaica, 37.) Wiesbaden: Harrassowitz, 93-102.

Johanson, Lars. (1993c). Typen türkischer Kausalsatzverbindungen. *Journal of Turkology* 2, 213-267.

Johanson, Lars. (1995a). On Turkic converb clauses. In: Haspelmath, Martin & König, Ekkehard (eds.) *Converbs in cross-linguistic perspective. Structure and meaning of adverbial verb forms – adverbial participles, gerunds* (Empirical approaches to language typology, 13), Berlin, New York: Mouton de Gruyter, 313-347.

Johanson, Lars. (1995b). Wie entsteht ein türkisches Wort? Kellner, Barbara and Marek Stachowski (eds.) *Laut- und Wortgeschichte der Türksprachen* (Turcologica, 26), Wiesbaden: Harrassowitz , 97-121.

Johanson, Lars. (1996a). Kopierte Satzjunktoren im Türkischen. *Sprachtypologie und Universalienforschung* 49, 1-11.

Johanson, Lars. (1996b). Altaische Postterminalia. *Acta Orientalia Hungarica* 49, 257-276.

Johanson, Lars. (1997). Kopien russischer Konjunktionen in türkischen Sprachen. In: Huber, Dieter and Erika Worbs (eds.) *Ars transferendi. Sprache, Übersetzung, Interkulturalität*, Frankfurt: Peter Lang, 115-121.

Johanson, Lars. (1998). Code-copying in Irano-Turkic. *Language Sciences* 20, 325-337.

Johanson, Lars. (1999a). Frame-changing code-copying in immigrant varieties. In: Extra, Guus and Ludo Verhoeven (eds.) *Bilingualism and migration*, Berlin, New York: Mouton de Gruyter, 247-260.

Johanson, Lars. (1999b). The dynamics of code-copying in language encounters. Brendemoen, Bernt et al. (eds.) *Language encounters across time and space*, Oslo: Novus Press, 37-62.

Johanson, Lars. (1999c). Cognates and copies in Altaic verb derivation. Menges, Karl H. and Nelly Naumann (eds.) *Language and literature—Japanese and the other Altaic languages. Studies in honour of Roy Andrew Miller on his 75th birthday,* Wiesbaden: Harrassowitz, 1-13.

Johanson, Lars. (In print). Code-copying and linguistic convergence. Esch, Edith, Mari Jones and Agnes Lam (eds.) *Linguistic convergence*, London: Arnold Publishers.

Johanson, Lars and Éva Á. Csató (eds.) (1998). *The Turkic languages*, London, New York: Routledge.

Epenthesis-Driven Harmony in Turkish

ABIGAIL R. KAUN
Yale University

0. Introduction

Describing one particular language requires only a subset of the Optimality Theoretic (Prince and Smolensky 1993) constraints necessary for the characterization of all phonological phenomena in all languages. Nonetheless, it is generally assumed that all grammars contain the same (universal) constraints, and differ only with respect to how those constraints are ranked relative to one another. The model therefore entails that individual grammars contain numerous latent or inactive constraints. This paper provides evidence that constraints which never play an active role in determining optimal outputs in native vocabulary are nonetheless available to speakers and may be invoked when novel phonological contexts are encountered, as in the case of loanwords. Several studies have reached similar conclusions for Tagalog (Ross 1996), Finnish (Ringen and Heinämäki 1999) and Tuvan (Harrison, this volume).

Words borrowed into Turkish from foreign sources typically undergo epenthesis to break up an initial consonant cluster. For instance, *steno* is realized as *siteno*. These epenthetic vowels are always high, however their backness and rounding are contextually determined, as noted in Yavas (1980) as well as Clements and Sezer (1982). This paper reports on the findings of an experiment conducted to determine the range of harmonic patterning of such epenthetic vowels. The results indicate that as described in the earlier studies, regressive harmony does target epenthetic vowels. The harmony "facts" vary considerably from speaker to speaker, however, and within speakers it is frequently the case that more than one harmony pattern is judged to be acceptable for a given word. I present an account of the results of this experiment focusing in particular on cross-linguistic evidence for the existence of the relevant inactive constraints that subjects must recruit to accommodate the novel contexts introduced in loanwords.

1. Patterns of regressive rounding harmony

Regressive harmony targeting epenthetic vowels in loanwords is described in Yavas (1980) and Clements and Sezer (1982). Both demonstrate that the backness of an epenthetic vowel is contextually determined. Vowels as well as velar and lateral consonants serve as backness harmony triggers. Similarly, these studies also document regressive rounding harmony targeting epenthetic vowels. The focus of this paper is limited to the phenomenon of regressive rounding harmony (hereafter RRH).

Both Yavas (1980) and Clements and Sezer (1982) report RRH targeting epenthetic vowels, however the patterns described are somewhat different. For both studies, high rounded vowels consistently trigger rounding harmony, as shown in the examples in (1):

(1)   High vowels consistently trigger RRH
    a. fülüt   'flute'   (Yavas)
    b. gurup   'group'   (Yavas, Clements and Sezer)
    c. Purusya   'Prussia'   (Clements and Sezer)

Different patterns are reported when the potential trigger is non-high. Yavas' data and discussion indicate that RRH is never triggered by a non-high vowel. Examples are shown in (2):

(2)   Yavas: Non-high vowels do not trigger RRH
    a. bironz (*buronz) 'bronze'
    b. filört (*fülört) 'flirt'

Clements and Sezer's data include some words in which a non-high vowel triggers RRH (3a-b), others in which a non-high vowel does not trigger RRH (3c-d), and one word (3e) for which two forms are listed, one containing a rounded epenthetic vowel and the other containing an unrounded epenthetic vowel:

(3)   Clements and Sezer: Non-high vowels sometimes trigger RRH
    a. purova   'test'
    b. purotesto   'protest'
    c. biroč   'brooch'
    d. kiroki   'sketch'
    e. burom / birom   'bromide'

The experiment described in the present paper was originally designed to resolve this discrepancy. As will be shown, however, both of the patterns cited above were documented in the experiment described here. Additional patterns were also observed.

1.1.   Experimental procedure

The data were collected in the following manner. Nine native speakers of Turkish, ranging in age from 18-35 years, were interviewed. All subjects had been raised in urban settings including Ankara, Istanbul and Izmir. Each subject was presented with a list of 107 words of foreign origin taken from Özgüler (1989). Each word contained an initial two-consonant cluster. Subjects were asked to determine the appropriate epenthetic vowel quality (or qualities). The task presented little difficulty for the subjects, and responses were supplied with what appeared to be a high degree of confidence. Certain subjects reported two vowel qualities to be appropriate for certain words.

The results of this survey were all consistent with Yavas (1980) and Clements and Sezer (1982) in one respect: High rounded vowels always triggered RRH. When the potential trigger was non-high, however, considerable subject-to-subject variation was observed.

1.2.   Results

Six patterns of RRH emerged from the results of this experiment. Of these, three involve intra-speaker variation, i.e., patterns in which certain vowels trigger RRH only optionally. The six patterns are given below in (4). In this table, solid lines enclose consistent RRH triggers, while dashed lines enclose optional triggers. Un-enclosed vowels never trigger harmony for the pattern in question. The number of subjects instantiating each pattern is indicated in parentheses:

(4)   Six rounding harmony patterns

Group A	(1)	$\boxed{\text{ü}\qquad\text{u}}$	ö	o
Group B	(1)	$\boxed{\text{ü}\qquad\text{u}}$	$\underline{\overline{\text{ö}}}$	o
Group C	(2)	$\boxed{\text{ü}\qquad\text{u}\qquad\text{ö}}$		o
Group D	(1)	$\boxed{\text{ü}\qquad\text{u}}$	$\underline{\overline{\text{ö}\qquad\text{o}}}$	
Group E	(2)	$\boxed{\text{ü}\qquad\text{u}\qquad\text{ö}}$	$\underline{\overline{\text{o}}}$	
Group F	(2)	$\boxed{\text{ü}\qquad\text{u}\qquad\text{ö}\qquad\text{o}}$		

So in group B, for example, RRH is consistently triggered by the high vowels, is optionally triggered by the non-high front vowel [ö], and is never triggered by the non-high back vowel [o]. (5) shows some of the words used in the experiment:

(5)   Sample Words

	'flute'	'blouse'	'diploma'	'block'
Group A	fül̲üt	bul̲uz	bir̲öve	bıl̲ok
Group B	fül̲üt	bul̲uz	bir̲öve/bür̲öve	bıl̲ok
Group C	fül̲üt	bul̲uz	bür̲öve	bıl̲ok
Group D	fül̲üt	bul̲uz	bir̲öve/bür̲öve	bıl̲ok/bul̲ok
Group E	fül̲üt	bul̲uz	bür̲öve	bıl̲ok/bul̲ok
Group F	fül̲üt	bul̲uz	bür̲öve	bul̲ok

From these patterns we may draw two conclusions. First, high vowels are clearly more popular RRH triggers than non-high vowels. Second, of the non-high vowels, the front vowel [ö] is more likely to trigger RRH than the back vowel [o] is. These asymmetries are implicational: Whenever a non-high vowel may trigger RRH, high vowels may also trigger RRH. Similarly, whenever a back vowel serves as a RRH trigger, the corresponding front vowel does so as well.

These generalizations cannot be predicted on the basis of the progressive rounding harmony pattern of native Turkish words. In native rounding harmony, vowel height does play a role in determining the applicability of harmony, however it is the target whose height is relevant, rather than the trigger. Moreover, backness never serves to restrict the applicability of native rounding harmony. If we look beyond standard Turkish, however, we observe patterns of rounding harmony rather similar to those obtained in the experiment described here.

2. Rounding harmony beyond Turkish

In Kaun (1994), I showed that clear patterns emerge when the typology of rounding harmony is submitted to cross-linguistic examination. That study attempted to explain the fact that most attested rounding harmony systems cannot be described by a simple statement such as "a vowel is rounded when it occurs in the vicinity of a rounded vowel." Rather, in nearly all documented cases, conditions referring to the height and/or backness of the participating vowels are imposed on the application of harmony. These include the following:

(6)   Cross-linguistically observed conditions on rounding harmony
    a. The target must be high.
    b. The trigger must be front.
    c. The trigger and target must agree in height.

In Nawuri (Casali 1993) and Tuvan (Harrison, this volume), for example, rounding harmony is subject to condition (6a). In Yakut (Krueger 1962), rounding harmony applies as long as condition (6a) or condition (6c) is met. In Tsou (Hsu 1993), rounding harmony obtains as long as conditions (6a) and (6c) are met, while Karakalpak (Menges 1947) requires either (6a) or (6b). Other combinations of these conditions describe other attested rounding harmony patterns.

2.1.   A formal account of the typology

We may formalize the general tendencies described above as Optimality Theoretic constraints (Prince and Smolensky 1993). Elsewhere (Kaun 1994, forthcoming), I discuss the phonetic motivation of these constraints. In the present paper, however, the constraints are posited with at most only general remarks regarding their functional underpinnings.

First, many languages exhibit rounding harmony. Following Smolensky (1993) and others, I will assume that the phenomenon of harmony is encoded grammatically by means of constraints which call for the alignment of the harmonic feature with the edge of some domain, typically the prosodic word. For rounding harmony, then, two general alignment constraints will be posited:

(7)   General alignment constraints for rounding harmony

    ALIGN-R([RD])   *The feature [round] is aligned with the right edge of the prosodic word.*

    ALIGN-L([RD])   *The feature [round] is aligned with the left edge of the prosodic word.*

We noted above that front rounded vowels are preferred over back rounded vowels as harmony triggers. This suggests the need for the more specific constraints listed in (8):

(8)   Specific alignment constraints for rounding harmony

    ALIGN-R([RD]/[-BK])   *The feature [round], when co-occurring with the feature [-back], is aligned with the right edge of the prosodic word.*

    ALIGN-L([RD] /[-BK])   *The feature [round], when co-occurring with the feature [-back], is aligned with the left edge of the prosodic word.*

The tendency for high vowels, rather than non-high vowels, to be targeted by rounding harmony may be attributed to the general markedness dispreference for non-high rounded vowels. We encode this dispreference formally in a constraint banning the combination of [round] and [low]:

(9)   *ROLO   *The combination [round]+[low] is dispreferred.*

Finally, we noted above that rounding harmony is favored in an environment in which the trigger and target agree in height. That is, cross-height harmony is disfavored. Citing work by Goldstein (1991), which demonstrates that the lip-rounding gesture associated with high vowels differs substantively from that associated with non-high vowels, I have attributed this phenomenon to a constraint on *gestural uniformity*. Gestural uniformity constraints state that in the phonological representation, a single instance of a feature should correspond to a uniform execution mechanism for the corresponding gesture. A single instance of [round] associated to vowels of differing heights will incur a violation of the following gestural uniformity constraint:

(10)   Gestural uniformity
    GESTUNI[RD]   *The feature [round] should be realized with a uniform mechanism of articulation.*

2.2.   Native Turkish rounding harmony

For native Turkish rounding harmony, only a subset of the constraints proposed in section 2.1 plays a decisive role in determining surface forms. High vowels undergo rounding harmony triggered by both high and non-high vowels,

as shown in (11). Non-high vowels do not undergo rounding harmony, as shown in (12). Note that these suffixes (as well as epenthetic vowels) also undergo backness harmony. No formal account of backness harmony is proposed here.

(11)   High vowels undergo rounding harmony
      pul   pul-<u>u</u>m   'stamp-1.singular'
      kol   kol-<u>u</u>m   'arm-1.singular'
      süs   süs-<u>ü</u>m   'ornament-1.singular'
      söz   söz-<u>ü</u>m   'word-1.singular'

(12)   Non-high vowels do not undergo rounding harmony
      pul   pul-l<u>a</u>r   'stamp-plural'
                        (*pul-l<u>o</u>r)
      kol   kol-l<u>a</u>r   'arm-plural'
                        (*kol-l<u>o</u>r)
      süs   süs-l<u>e</u>r   'ornament-plural'
                        (*süs-l<u>ö</u>r)
      söz   söz-l<u>e</u>r   'word-plural'
                        (*söz-l<u>ö</u>r)

The native pattern of rounding harmony is progressive, and targets only high vowels. This pattern reflects the interaction of the two constraints in (13), where (a) outranks (b):

(13)   Active constraints for Turkish
      a.   *ROLO
      b.   ALIGN-R([RD])

The remaining constraints play no active role in native Turkish: No regressive harmony is exhibited (14a), front vowels are not preferred over back vowels as harmony triggers (14b-c), and cross-height harmony is freely permitted (14d):

(14)   Inactive constraints for Turkish
      a.   ALIGN-L([RD])
      b.   ALIGN-R([RD]/[-BK])
      c.   ALIGN-L([RD]/[-BK])
      d.   GESTUNI[RD]

The typological facts suggest that rounding harmony can be characterized cross-linguistically with a collection of potentially interacting constraints. A particular rounding harmony system will typically be describable with just a subset of those constraints. My interest lies in the remaining constraints; those which play no active role in determining the native patterns of the language in question. While the inactive constraints listed in (14) are not actively invoked in the general native pattern of rounding harmony, they are indeed available to Turkish speakers, and are recruited when a novel phonological context is

presented, as in the case of loanwords. This claim is substantiated in the analysis proposed in section 3.

3. Analysis

The loanword context described here sets up the environment for *regressive* harmony, rather than the natively observed progressive harmony. This domain of borrowing thus constitutes a laboratory for observing the accessibility of the (natively inactive) left-alignment constraints. Moreover, if regressive harmony and progressive harmony result from distinct alignment constraints, we might expect to find that the constraint interactions exhibited in RRH are different from those associated with native progressive harmony. Since epenthetic vowels are always high in Turkish, the cluster-initial loans do not allow us to observe the interaction of the feature co-occurrence constraint *ROLO with the regressive harmony constraints. However, Turkish has borrowed heavily from French, a language possessing both front and back rounded vowels. Thus, the typologically observed front-back trigger asymmetry could in principle interact with RRH in Turkish. By the same token, it is in principle possible for gestural uniformity effects to be observed under Turkish RRH. Both of these possibilities are in fact borne out.

Regressive rounding harmony clearly exists. All subjects exhibited RRH when the potential trigger was a high vowel. From this we conclude that while not natively active, left-alignment constraints are indeed accessible to speakers. This observation might not in itself constitute conclusive proof that grammars contain inactive constraints that can be accessed when the need arises. It could be the case that, contrary to the analysis proposed here, rounding harmony is encoded in a single "bi-directional" constraint. Under this view, the loanword patterns would be the expected consequence of the fact that loanwords, but not native words, introduce a potential RRH environment. I will show that the interaction of RRH with backness and gestural uniformity demonstrates that latent constraints do exist and that speakers access them readily.

3.1. The accessibility of inactive constraints

As described in section 1, subjects differ with respect to the application of RRH when the potential trigger is a non-high vowel. For one group (group F), non-high rounded vowels consistently trigger RRH, whereas for another group (group A), non-high rounded vowels consistently *fail* to trigger RRH. This difference reflects two distinct rankings of the constraint which favors RRH (left-alignment) vis-à-vis that which expresses a dispreference for cross-height harmony (gestural uniformity). Group A assigns a higher rank to gestural uniformity, whereas group F assigns a higher priority to left-alignment:

(15)   Permitting vs. prohibiting cross-height harmony
        Group A   GESTUNI[RD] >>   ALIGN-L([RD])
        Group F   ALIGN-L([RD])>>   GESTUNI[RD]

Group D shows a "compromise" pattern—one in which the non-high vowels optionally trigger RRH. We return to optionality in section 3.2.

The remaining patterns exhibit asymmetric treatment of front versus back triggers, suggesting a higher ranking of the backness-specific left-alignment constraint relative to the more general left-alignment constraint. In the group C pattern, [ö] consistently triggers RRH, whereas [o] consistently does not. For speakers exhibiting this pattern, we may suppose that the general left-alignment constraint is ranked low relative to gestural uniformity, while that favoring front vowel-triggered harmony is ranked higher:

(16)   The front-back trigger asymmetry
        Group C   Align-L([rd]/[-bk]) >> GestUni[rd] >> Align-L([rd])

Recall that native Turkish does not, in itself, provide evidence for the existence of a special constraint calling for the alignment of [round] when it co-occurs with [-back]. The behavior of group C (and others) in the present experiment demonstrates that the grammar of Turkish must include both sets of alignment constraints despite the fact that native patterns supply positive evidence for only the more general [round]-alignment.

The remaining patterns involve optionality and are discussed in section 3.2.

3.2.   Optionality as *indeterminate ranking*

Let us begin with group B speakers, for whom the non-high front rounded vowel [ö] optionally triggers RRH. To account for this pattern, suppose that two distinct constraint hierarchies are available to this subset of speakers: one in which gestural uniformity outranks left-alignment, and the other in which left-alignment outranks gestural uniformity. Expressed formally, we may posit an *indeterminate ranking* of GESTUNI[RD] and ALIGN-L([RD]/[-BK]) and, as shown in (17). Indeterminate ranking is denoted with the symbol (><):

(17)   Indeterminate ranking: group B
        ALIGN-L([RD]/[-BK]) >< GESTUNI[RD]

For group B speakers, the indeterminately ranked constraints in (17) must outrank the more general left-alignment constraint. The indeterminate ranking relation established in (17) yields a grammar in which two constraint hierarchies are possible. These are shown in (18):

(18) Two hierarchies for group B

	ALIGN-L([RD]/[-BK])	GESTUNI[RD]	ALIGN-L([RD])
→ **büröve**		*	
biröve	*!		*
bulok		*!	
→ **bɨlok**			*

	GESTUNI[RD]	ALIGN-L([RD]/[-BK])	ALIGN-L([RD])
büröve	*!		
→ **biröve**		*	*
bulok		*!	
→bɨlok			*

For group E speakers, for whom [ö] is a consistent RRH trigger, while [o] is an optional trigger, we may posit an indeterminate ranking of the general ALIGN-L([RD]) relative to gestural uniformity:

(19) Indeterminate ranking: group E
GESTUNI[RD] >< ALIGN-L([RD])

For this pattern, the more specific alignment constraint referring to front rounded vowels will rank higher than the two indeterminately ranked constraints, as given in (20). For the sake of brevity, tableaux are not included, however they would be analogous to those shown in (18).

(20) Two alternative hierarchies for group E
ALIGN-L([RD]/[-BK]) >> GESTUNI[RD] >> ALIGN-L([RD])
ALIGN-L([RD]/[-BK]) >> ALIGN-L([RD]) >> GESTUNI[RD]
Finally, let us consider group D speakers, for whom both [ö] and [o] optionally trigger RRH. To account for this pattern, we again posit an indeterminate ranking of general left-alignment and gestural uniformity. This pattern differs from the group E pattern however, in that the indeterminately ranked constraints will outrank the backness-specific left-alignment constraint, as given in (21):

(21) Two alternative hierarchies for group D
GESTUNI[RD] >> ALIGN-L([RD]) >> ALIGN-L([RD]/[-BK])
ALIGN-L([RD]) >> GESTUNI[RD] >> ALIGN-L([RD]/[-BK])

3.3. An alternative to indeterminate ranking

An alternate analysis would be simply to state that optionality results from certain constraints being unranked relative to one another. This approach is developed in Ringen and Heinämäki 1999 to account for variability in the choice of suffix vowels following disharmonic loanwords in Finnish. An account involving unranked constraints works well to characterize the group E pattern, in which [o] triggers RRH only optionally. Consider the tableau in (22), in which forms containing both a rounded and an unrounded epenthetic vowel are selected when the potential trigger is [o]:

(22)   Group E: GESTUNI[RD] unranked with respect to ALIGN-L([RD])

	ALIGN-L([RD]/[-BK])	GESTUNI[RD]	ALIGN-L([RD])
→ büröve		*	
biröve	*!		*
→ bulok		*	
→ bilok			*

For the remaining groups exhibiting optionality, however, an analogous analysis is not workable. This is due to the fact that in the event of a tie, the decision will be left to the next constraint in the hierarchy. Consider the tableau in (23), which is meant to characterize the group B pattern. Recall that in group B, [ö] optionally triggers RRH, while [o] consistently fails to trigger RRH. As shown, the hierarchy in which the two relevant constraints are unranked fails to select one of the two acceptable forms, as indicated with (←):

(23)   Group B: GESTUNI[RD] unranked with respect to ALIGN-L([RD]/[-BK])

	ALIGN-L([RD]/[-BK])	GESTUNI[RD]	ALIGN-L([RD])
→ büröve		*	
← biröve	*		*!
bulok		*!	
→ bilok			*

When the unranked constraints yield a tie, the decision goes to the more general alignment constraint which follows in the hierarchy, and no optionality is predicted. The unranked analysis confronts the same problem when trying to characterize group D. Since two constraints conspire to yield harmony and only one has the effect of blocking harmony, the unranked analysis offers no means of characterizing optionality when the potential trigger is a front vowel.

Of course, if an additional anti-harmony constraint were to be posited above the general left-alignment constraint, an unranked constraint analysis of optionality could be maintained. For the data under analysis in the present paper, however, there is another sense in which the unranked analysis is less plausible than the indeterminate ranking analysis proposed here. For words with two "correct" forms (e.g., bulok and bilok 'block' for group E speakers), we would expect both forms to be equally acceptable *at all times*. This prediction is not consistent with the subjects' performance, however.

When subjects from groups B, D, and E performed the task in the experimental setting, they often supplied only one of the possible forms for a given word. That is, for certain tokens they supplied harmonic forms, for others they supplied non-harmonic forms, and for still others they supplied two forms. Excerpts from one data sheet of a group E subject are given in (24):

(24)   Partial results: one group E speaker
        a. /plörit/   [pülörit]
        b. /bröve/   [biröve]
        c. /blöf/   [bülöf] or [bilöf]

We can model this type of behavior by positing indeterminate ranking as proposed above. On the occasion when (24a) was produced, the subject applied the ranking in which left-alignment outranks gestural uniformity. By contrast, when (24b) was produced, the alternate ranking was used. When two responses were given, the speaker made use of each hierarchy in turn. If the relevant constraints were analyzed as being unranked relative to one another, we would expect consistently to encounter responses like that given in (24c), in which both of the tied candidates are judged to be acceptable.

4. Discussion

In this paper, I have argued that the facts of RRH in Turkish loanwords demonstrate that speakers access inactive constraints when confronted with a phonological environment that does not arise in the structures native to their language. The cross-linguistic patterning of the phenomenon under investigation suggests the existence of a collection of potentially interacting constraints, only two of which are decisive in the native structures of Turkish. The remaining constraints, despite their inertness in the native phonology, must nonetheless be present in the grammar of Turkish speakers. These results support the standard Optimality Theoretic claim that languages share a common set of constraints and distinguish themselves idiosyncratically only with respect to how those constraints are ranked.

I have also sketched a formal account of optionality consistent with the experimental results described here. As Ross (1996) has pointed out, it is not surprising that speakers should exhibit variability with respect to the ranking of inactive constraints. Such constraints come into play only when non-native phonological contexts are encountered—anomalous phonological circumstances for which no conventions have as yet been firmly established within the speech community.

References

Casali, Roderic. (1993). Cross-height vowel harmony in Ghanaian. Ms., UCLA.

Clements, George N. and Engin Sezer. (1982). Vowel and consonant disharmony in Turkish. In Harry van der Hulst and Norval Smith (eds.), *Structure of phonological representations* (Vol. II), pp. 213-255. Dordrecht, Holland: Foris Publications.

Goldstein, Louis. (1991). Lip rounding as side contact. *Actes du XIIème congrès international des sciences phonetiques, Vol I/5: Sessions plenieres et semiplenieres.* 19-24 August, Aix-en-Provence, France.

Harrison, K. David. (this volume). Tuvan reduplication and harmony.

Hsu, Chai-Shune. (1993). Tsou phonology. Ms., UCLA.

Kaun, Abigail R. (1994). An Optimality Theoretic account of rounding harmony typology. *The Proceedings of the 13th West Coast Conference on Formal Linguistics.*

Kaun, Abigail R. (forthcoming). Rounding harmony. In Bruce Hayes, Robert Kirchner and Donca Steriade (eds.) *Phonetics in Phonology.*

Krueger, John. (1962). *Yakut manual*. Uralic and Altaic Series, Vol. 21. Bloomington, Indiana: Indiana University Publications.

Özgüler, Ali. (1989). *Türçedemizdeki yabanci kelimeler sözlük*.

Prince, Alan and Paul Smolensky. (1993). Optimality theory: Constraint interaction in generative grammar. *Technical Report #2 of the Rutgers Center for Cognitive Science*, Rutgers University.

Menges, Karl. (1968). *The Turkic languages and peoples: An introduction to Turkic studies*. Wiesbaden: Otto Harrassowitz.

Ross, Kie. (1996). Floating phonotactics: Variability in infixation and reduplication of Tagalog loanwords. MA. Thesis, UCLA.

Ringen, Catherine, and Orvokki Heinämäki. (1999). Variation in Finnish vowel harmony: An OT account. *Natural Language and Linguistic Theory*, **17** (2), 303-337.

Smolensky, Paul. (1993). Harmony, markedness, and phonological activity. Paper delivered at the Rutgers Optimality Workshop 1.

Yavas, Mehmet. (1980). Borrowing and its implications for Turkish phonology. Ph.D. dissertation, University of Kansas.

Language policy and reforms of Uygur and Kazak writing systems in China *

MINGLANG ZHOU
University of Colorado, Boulder

1. Introduction

China officially has fifty-five ethnic minority nationalities (EMN's) with a combined population of about 91 million (1990 census). Members of this population speak more than eighty languages, most of which do not have written forms. In the last forty-plus years, under the auspices of the Chinese government and the guidance of its EMN language policy, thirty versions of writing systems have been created for some oral languages, and nineteen existing writing systems have been reformed or revised (Zhou 1999a). These new and revised writing systems have had profound impacts on the education, literary traditions, culture, and daily life of the 91 million EMN people in China.

Among the nineteen reformed and/or revised writing forms are the writing systems of Uygur and Kazak, two Turkic languages, that underwent a major reform from the Arabic script to the Latin alphabet in the 1960s, and a second major reform from the Latin Alphabet back to the revised Arabic script in the early 1980s, in addition to minor revisions of the Arabic alphabet carried out during the early 1950s and the abortive introduction of a Cyrillic writing system in the middle of the 1950s. Detailed information on these reforms does not seem to be available to linguists and writing system specialists outside China (cf. Coulmas 1989, Coulmas 1996, Daniels and Bright 1996, Johnson and Csato2 1998).

This paper briefly reviews the Chinese government's EMN policies, which dictate its language policy; describes the processes by which the old writing systems were revised and new ones were introduced; investigates the impact that these reforms have had on more than seven million Uygur speakers and over one million Kazak speakers (1990 census) in China; and examines the significance of the Chinese government's language policy for language policy makers and researchers in China and worldwide.

2. EMN policies and language policy in China

Since 1949, the Chinese Communist Party's (CCP) EMN policies have been theoretically guided by Marxist-Leninist views of nation and nationality, and practically modeled after those of the former Soviet Union (cf. Dreyer 1976, Dreyer 1997, Mackerras 1994). In nationality classification, for example, the Chinese government adopts Stalin's definition of nationality as being based on common language, common territory, common economic life, and common culture (see AOPROSNAC 1996), but it has never applied this definition dogmatically, as Zhou Enlai, premier of China 1949-1976, pointed out (cf. Liu and Zhang). In China, nationality (minzu) is instead defined by one or more of those common features in either contemporary or historical terms. Particularly influential has been that nationalities rise in the development to capitalism and converge in the development to communism (cf. Stalin 1975). The key issue for the CCP is the pace of the development between capitalism and communism. Is it a long-term development, as has been the official view since 1979, or a short-term development, as was believed at the time of the Great Leap Forward and throughout the Cultural Revolution? Both views consider nationalism a stage to be transcended in the international unification

of the proletariat; the question is whether, in the immediate present, it should be tolerated, discouraged, or opposed. The struggle between these two views has influenced the CCP's EMN policies, which in turn have swayed the Chinese government's language policy for EMN languages and resulted in the frequent reforms of Uygur and Kazak writing systems.

The right for EMN peoples to use and develop their native languages and writing systems has been guaranteed constitutionally in China, since the provisional of 1949, with reaffirmations in 1954, 1975, 1978, 1982, and 1999. However, there has been a gap between the constitutional guarantee and Chinese practice.

In the early 1950s, the long-term development view was dominant within the CCP and the Chinese government. In 1951, the State Council established an EMN language and writing system research and guidance committee within the Ministry of Education (cf. EBCCNW 1989). In the following years, this committee sent teams of linguists to EMN communities to study EMN languages and to work with local EMN autonomous governments on creating writing systems for oral languages and on reforming existing writing systems.

The first sign of change came at the beginning of 1956. First, in January of that year, the Central Committee of the CCP set a two-to-three year timetable for creating and reforming EMN writing systems (cf. EBCCNW 1989). Second, in March of the same year, the State Council centralized authority over EMN writing systems, making the EMN Language Institute of the Chinese Science Academy responsible for creating writing systems, the State Nationality Affairs Commission responsible for checking and approving them, the Ministry of Education responsible for experimentation and implementation, the Ministry of Culture responsible for translation and publication in EMN languages, and the Central Institute for Nationalities responsible for training EMN linguists and language educators. These regulations deprived local EMN autonomous governments of authority in EMN language affairs.

As late as the beginning of 1956, the Chinese government's language policy for EMN languages was still independent of its language policy for Chinese. An instruction by the CCP Central Committee in January 1956 clearly stated that Putonghua (PTH, the official national language based on Mandarin) was to be promoted among the Hans only (see PROSCLS 1996). And an instruction issued by the State Council in February 1956 stated that in EMN autonomous regions PTH was to be promoted only within Han communities and in Chinese language classes for EMN students (see PROSCLS 1996). But in April 1956, Mao Zedong, in a speech " On Ten Relationships", expressed the short-term development view that both Han chauvinistic nationalism and EMN nationalism should be fought against as obstacles to the unity of nationalities under socialism. In July 1957, Zhou Enlai echoed this adjuration. A further turning point came in September of the same year, when at a CCP Central Committee meeting, Deng Xiaoping, then general secretary of the CCP, emphasized that in the Antirightist Struggle the focus was on the struggle against local nationalism in EMN regions (EBCCNW 1989:80). Finally, in 1964, *Red Flag*, the journal of the CCP, published an article "On current nation/nationality issues and class struggle in our country", which asserts that nation/nationality issues are class issues. This assertion had great implications: anyone who raises any issue not in favor of the rapid assimilation into the majority Han could be considered a class enemy.

When the short-term development view became dominant in 1958, implementation of language policy changed radically, though the constitutionally guaranteed EMN language right was still there. The change was first seen in the

convergence of the language policy for EMN languages and that for Chinese. After the Plan for the Phonetic Spelling of Chinese (in the Latin alphabet) was officially published in November 1957, the State Council made it clear that the Plan would play a significant role in the creation of writing systems for EMN languages and in EMN peoples' learning of PTH (see PROSCLS 1996). This role was spelled out by Zhou Enlai in January 1958: the Plan would serve as the common base for creating and reforming EMN writing systems which would use the Latin alphabet with pronunciations and usage similar to those in the Plan (Liu and Zhang 1994:190-1, Seybolt and Chiang 1979:236-7). From then on, the Chinese government's language policy for EMN languages became part of its language policy for Chinese. During the Cultural Revolution, the Chinese government's language policy was mostly ignored because of weakened governments at various levels or occasionally carried out radically in some regions because of some radicals in power.

At the end of 1978, during the third session of the Eleventh Central Committee of the CCP, the long-term development view finally began to win over the short-term development view. A day after the meeting, the State Nationality Affairs Commission and the Chinese Academy of Social Sciences held an academic forum focusing on two key issues in nation/nationality affairs: the timetable of nation/nationality convergence and the nature of nation/nationality issues. Consensus was reached that nation/nationality would exist for a long time, and that nation/nationality issues were essentially not class issues. This forum paved the way for a more pluralistic approach in the CCP's EMN policies and in the Chinese government's language policy for EMN languages. On July 15, 1980, *The People's Daily* carried an article, criticizing the 1964 *Red Flag* article and stating in public, for the first time in more than twenty years, that nation/nationality issues are not class issues. In the same year, the third national conference on EMN languages called for more respect of constitutionality in EMN language use and writing system reforms, and made long-term plans for EMN language research. Many ideas put forward during the 1979 and 1980 conferences were adopted in the revision of the PRC constitution in 1982 and in the drafting of the 1984 legislation on regional EMN autonomous governing, which specifically states that local EMN autonomous governments guarantee their people the right to use their native languages and develop their writing systems, and decide what EMN language(s) shall be adopted in regional government business.

3. Uygur and Kazak writing system reforms

Historically, Uygur and Kazak speakers have used several scripts for their respective languages, but by 1949 both had settled on slightly different Arabic scripts (cf. Benson and Svanberg 1998, Niu 1997). In the early 1950s, when the EMN language policy basically focused on improving existing writing systems and eradicating mass illiteracy, reform efforts were devoted to technical details and to matching between oral languages and written forms so that it would be easy for people to learn to read in the mass campaign against illiteracy. In 1952, revisions were made in the Arabic alphabet for Uygur. In the then current Arabic alphabet, the letter ﯤ was used to represent two vowels ([e], [i]) and one consonant ([j]), while the letter ﻭ was used to represent four vowels ([u],[y],[o], [ø]) and one consonant ([v/w]). To simplify the relationship between letters and phonemes, the letter ﯤ was replaced with ﯤٴ for [e], ﯨ for [i], and ﯤ for [j]. At the same time, the letter ﻭ was replaced with ﯗ for [u] and [y], ﯙ for [o] and [ø], and ﯞ for [v/w]. In 1954, minor

revisions were made in the Arabic alphabet where the letter ـۇ was changed to ـﯧ for [e] and the letter ـﯩ to ـﯨ for [i] (see Hu 1979, SOCWSRXUAR 1962). These revisions led to what is known as the old Uygur writing system, as shown in (1) below.

The revision of the Kazak writing system did not take place until 1954. There was a discussion among Kazak linguists, educators and publishers about the letter ۋ for [w] and the letter ى for [j]. Before the revision, the letter ۋ represented the consonant [w], while the letter cluster ۇۋ represented the glide [uw], and the cluster ىۋ stood for [yw], and the letter ى represented the consonant [j], whereas the letter cluster ىﻱ stood for [əj], and the cluster ﻱﻱ represented [ij]. After the revision, the letter ۋ represented [w] before or after vowel letters, while it stood for [uw] and [yw] after consonant letters. Similarly, the ى represented [j], when it appeared before or after vowel letters, but [əj] and [ij] after consonant letters (Geng 1989:32-37, SOCWSRXUAR 1962). These revisions ended in what is usually called the old Kazak writing system, as in (2) below.

Debates about the best scripts for Uygur and Kazak, however, continued in Xinjiang. Given what appeared to be a close relationship between China and the former Soviet Union at that time, the Cyrillic alphabet became a good candidate. The Uygur and Kazak communities had very different attitudes toward the Cyrillic alphabet. The Uygur communities used to have closer ties with the idea of a Pan-Turkestan and Pan-Turkism than the Kazak communities did (cf. XLJBZ 1993). For example, the Turkish-Islamic Eastern Turkestan Republic was established, with British and Islamic support, in south Xinjiang in 1933, but crashed by the Xinjiang government with Soviet assistance in about a year. The idea of the Eastern Turkestan continues to influence Uygur communities in some ways (cf. Mackerras 1994, XLJBZ 1993). On the other hand, the Kazak communities used to have closer ties with the Kazaks in the former Soviet Union, where eighty percent of all Kazaks live. There was relatively free movement of Kazaks across the border. Soviet publications in Kazak (in the Cyrillic Alphabet) were imported and used in Chinese Kazak schools in the early and middle 1950s (cf. Wang 1990). Given the two contrasting situations, the Uygur communities appeared to be much less enthusiastic about, if not hostile to, the Cyrillic alphabet than the Kazak communities were. In 1956, during the first Xinjiang Conference on EMN languages, a decision was made to adopt Cyrillic alphabets to replace the existing Arabic alphabets for Uygur, Kazak, and several other EMN languages in the next ten years. The ten-year plan might have been halfhearted, as the leadership of the CCP was suspicious of the intentions of the leadership of the former Soviet Union (cf. Benson and Svanberg 1998). It was, of course, later interrupted by the worsening relationship between China and the former Soviet Union. Most important of all, by late 1957, the short-term development view had become dominant and the Plan for the Phonetic Spelling of Chinese had been adopted by the central government. Therefore, the scheme to replace Arabic alphabets with Cyrillic ones for Uygur and Kazak was fortunately abandoned before it was put into practice in Uygur and Kazak communities in Xinjiang.

Soon after the Plan and Zhou Enlai's speech were published in 1958, Uygur and Kazak linguists and educators dutifully began discussing Latin alphabets for Uygur and Kazak. However, this time the atmosphere of the discussions was completely different. In the middle of December 1957, an expanded conference of the Xinjiang CCP Committee convened, which lasted over four months. At the conference, Seypidin Azizi, then governor of Xinjiang and secretary of the Xinjiang

CCP Committee (under the first secretary, Wang Enmou) made a speech opposing local (EMN) nationalism. A fierce struggle was seen at the four-month conference-- so fierce--that Seypidin was accused of being soft on local nationalism, and had to make self-criticisms during the conference to keep his position. Many EMN CCP officials, together with some Han CCP officials, were removed or demoted from their posts for their local nationalism or tolerance of local nationalism. Four months later, the conference of the Xinjiang CCP concluded, on April 28 1958, with a resolution on opposing and overcoming local (EMN) nationalism in Xinjiang.

How the short-term development view influenced discussions on writing system reforms for Uygur, Kazak, and other EMN languages is best seen in Seypidin Azizi's speech at the second conference on Xinjiang's EMN languages, held in December 1959 (cf. Azizi 1960). In his speech concerning nation/nationality and language, Seypidin Azizi first elaborated on Stalin's notion on the rise and convergence of nation/nationality, talked about the necessity of national convergence or assimilation under socialism, clarified the role of language as a tool for class and social struggle, and stressed the importance of fighting against bourgeois nationalism in language and script reforms, as he concluded:

> To oppose the adoption of the new writing systems based on the Plan for the Phonetic Spelling of Chinese and to oppose the development of EMN languages along with Chinese is to oppose socialism and communism of the Chinese peoples, and to oppose the creation or reform of writing systems for EMN languages based on the Plan is to oppose the unity of all Chinese nationalities and the unification of the motherland. (Azizi 1960:15)

Clearly any discussion on writing system reforms beyond the Plan was very risky in that situation at that time. Not surprisingly, the language conference, which was claimed to have overcome reactionary nationalist thoughts on EMN language and writing system development in Xinjiang, adopted Latin alphabets as the new writing systems for Uygur and Kazak and planned to utilize the new writing systems exclusively within three to five years.

To justify the adoption of the new writing systems, three major technical reasons were given in addition to a number of political and social justifications (SOCWSRXUAR 1962). First, it was claimed that the old writing systems could not fully represent the oral languages. For example, in Uygur, even after the revisions, the letter ﯗ stood for both [o] and [Ø], and the letter ﺋﯗ stood for both [u] and [y], resulting in one form for different meanings. A similar situation was also found in Kazak where the letter ﯞ represented [w], [uw], and [yw], and the letter ﯨ represented [j], [əj], and [ij]. Second, it was said that the old writing systems could not fully represent loan words from Chinese in the distinction between letters and letter combinations j [tɕ] and zh [tʂ], q [tɕ'] and ch [tʂ'], x [ɕ] and sh [ʂ], etc. Third, it was believed that the old writing systems were difficult for mass education. In the existing Arabic alphabets, each letter had two to six forms depending on its position in a word, which created difficulty in reading and writing, and there were too many diacritics, which created different meanings if missed or misplaced.

The scheme of the Latin alphabet for Uygur consisted of the twenty-six regular letters from the Plan and seven letters specially created to meet the needs of Uygur, allowed letter combinations ng, zh, ch, and sh intended for loan words from Chinese, and utilized the diacritic "," to represent syllable separation, as in (1).

(1) Comparison of the Old and New Writing Systems for Uygur with IPA

Latin:	Aa	Bb	Cc	Dd	Ee	Ff	Gg	Hh	Ii	Jj	Kk	Ll
Arabic:	ئا	ب	(تس) د		مُي	ڧ	گ	خ	ئى	ج	ك	ل
IPA:	a	b	(ts)	d	e	f	g	χ	i	dʒ	k'	l

Latin:	Mm	Nn	Oo	Pp	Qq	Rr	Ss	Tt	Uu	Vv	Ww
Arabic:	م	ن	ئۇ	پ	چ	ر	س	ت	ئۆ	(ۋ)	ۋ
IPA:	m	n	o	p'	tʃ'	r	s	t'	u	v	w

Latin:	Xx	Yy	Zz	Oǀoǀ	Ḥh	Ḳk	əə	Өө	Üü	Ƶƶ
Arabic:	ش	ي	ز	غ	ھ	ق	ئە	ئۇ	ئۈ	ژ
IPA:	ʃ	·j	z	ʁ	h	q'	ɛ	ø	y	ʒ

The scheme for the Latin alphabet for Kazak had the same twenty-six letters from the Plan and was complemented with six special letters to represent phonetic features in Kazak, as in (2). The alphabet also allowed the letter combinations *ng*, *zh*, *ch*, and *sh* for loans from Chinese.

(2) Comparison of the Old and New Writing Systems for Kazak with IPA

Latin:	Aa	Bb	Cc	Dd	Ee	Ff	Gg	Hh	Ii	Jj	Kk	Ll
Arabic:	ا	ب	(تس) د		ى	ڧ	گ	ح	ىٴ	ج	ك	ل
IPA:	a	b	(ts)	d	e	f	g	χ	i	dʒ	k'	l

Latin:	Mm	Nn	Oo	Pp	Qq	Rr	Ss	Tt	Uu	Vv	Ww	Xx
Arabic:	م	ن	و	پ	چ	ر	س	ت	ۇ	ۋ	ۋ	ش
IPA:	m	n	o	p'	tʃ'	r	s	t'	u	v	w	ʃ

Latin:	Yy	Zz	Oǀoǀ	Ḥh	Ḳk	əə	Өө	Üü	Êê
Arabic:	ي	ر	ع	ھ	ق	اٴ	ۇ	ۇٴ	ه
IPA:	j	z	ʁ	h	q'	ɛ	ø	y	e

After several months' preparation, the new writing systems were exclusively used in the first grade in elementary schools, and experimentally used in some other grades in elementary and secondary schools in the fall of 1960. The new writing systems were more and more extensively used in schools in the following years. In 1962, newspapers began to be published in the new script for Uygur. In 1964, the People's Congress of Xinjiang passed a resolution to exclusively utilize the new writing systems in the whole Xinjiang beginning on New Year's Day 1965. However, the implementation of the new writing systems was interrupted by the

Great Culture Revolution, which swept the whole country in 1966 and did not end until 1976. During the Cultural Revolution, all levels of government within Xinjiang were sometimes paralyzed, and most of the times occupied with more urgent "revolutionary tasks". The progress for exclusive utilization of the new writing systems was very slow in society at large. For example, the Kazak edition of *The Xinjiang Daily* did not appear in the Latin script until 1974. In private situations, older people still used the Arabic scripts in writing. However, the new writing systems were exclusively utilized from elementary schools to colleges in Xinjiang as well as in government business.

After the Culture Revolution ended in 1976, discussions on the different scripts and the disaster of two different writing systems being used simultaneously for one language were resumed in Xinjiang. Three years later, in November 1979, the second session of the Fifth People's Congress of Xinjiang took a bold unprecedented step toward reversing the reform: the new writing systems should continue to be fully implemented, but the old writing systems should be used alongside the new ones. Of course, this decision did not satisfy the Uygur and Kazak communities, nor did it stop discussions on the topic. Instead, it encouraged citizen discussions on the issue in Xinjiang, discussions that eventually led to a more complete reversal.

After the CCP's EMN policies became more and more clearly pluralistic, and the State Council set a precedent by approving a standardized Yi writing system in a Siniform script instead of the Latin script, the seventeenth session of the Fifth People's Congress of Xinjiang passed a resolution, on September 13, 1982, that called for exclusive utilization of the old Arabic writing systems at the beginning of 1983, reserved the new Latin writing systems as pronunciation assistance systems, and encouraged further research on the new writing systems. The rationales for this decision, as the resolution stated (see *The Xinjiang Daily*, first edition, September 15, 1982), were that the practice (the use of the new systems) in the past few years had shown that the time was not ripe for promoting the new writing systems, and that the simultaneous use of both the old one and the new one was not productive for the development of sciences, education and culture among the Uygur and Kazak peoples. The lack of a total rejection of the Latin alphabets in the decision might have been a political strategy that is intended to save face for the central government.

After twenty years of official use of the Latin writing systems and unofficial use of the old Arabic ones, there was a lot of confusion in orthography. A mixture of the previous versions was sometimes used, and nonstandard letters were also used. In 1983, the Language and Script Reform Commission of Xinjiang published further revised Arabic writing systems for Uygur and Kazak and required them to be used in publications, documents, and public signs beginning on January 1, 1984. In Uygur, the 1954 Arabic writing system used the letter ﯗ for both [o] and [ø], and the letter ﯗ for both [u] and [y]; in the newly revised Arabic writing system the letter ﯘ was added to represent [ø], and the letter ﯛ was added for [y] (cf. Niu 1997). In Kazak, the 1954 Arabic writing system did not have a letter for the nasal [ŋ]; the newly revised writing system added the letter ﯓ to represent [ŋ] (Geng 1989). These revisions of the Arabic writing systems may have brought them closer to the ideal of a one-to-one match between letters and phonemes in these two languages. Such a match was one of the motivations for the revisions and believed to facilitate mass education and literacy. The last revision appears to have put an end to the chaos of over thirty years of script reforms for Uygur and Kazak.

4. The impact of writing system reforms in Xinjiang

These radical reforms of the Uygur and Kazak writing systems, using completely different scripts, should have had serious consequences on Uygur and Kazak speakers, though the information is not readily available. Most sources published in China just mention that there was a chaos in which older Uygur and Kazak speakers used the old Arabic writing systems while middle-aged and young speakers used the new Latin ones (Niu 1997:155). Specifics of the chaos have not been published, since this is still a politically sensitive issue in China. However, for the sake of language policy makers in China and all over the world, it is worthwhile to explore what impact China's EMN language policy and the writing system reforms may have had on the Uygur and Kazak communities in Xinjiang.

First, we must ask how many Uygur and Kazak people have learned to read in the Latin scripts and have difficulty in reading the Arabic scripts. From their first use in schools in the fall of 1960 to their last use in schools in the spring of 1983, the Latin alphabets dominated Uygur and Kazak schools for twenty-three years. It is known that reading is learned through instruction in elementary schools (cf. Chomsky 1991, Snow 1991). The Uygurs and Kazaks who had their entire elementary education in the Latin scripts are those who were born between 1951, when people started to go to school a little bit late, and 1970, when people began schooling at about six or seven. Judging from the 1990 census (DPSSSB & EDSNAC 1994), this population is about two million among the Uygurs and 320,000 among the Kazaks. Of the two million Uygurs, 85 percent (1.7 million) have elementary education or more, while the rest (0.28 million) are illiterate or semi-literate (DPSSSB & EDSNAC 1994). Of the 320,000 Kazaks, about 95 percent (305,410) have elementary education or above, while the remaining 5 percent (15,480) are illiterate or semi-literate (DPSSSB & EDSNAC 1994). In addition, about 700,000 Uygurs and 135,000 Kazaks (born between 1971 and 1975) have their first two or more years' education in Latin scripts (DPSSSB & EDSNAC 1994). Thus, out of 7.2 million Uygurs, 2.4 million (33%) have been directly and significantly affected by the script reforms in their education, while out of 1.1 million Kazaks, around 400,000 (39%) have been directly and significantly influenced by the script reforms in their education.

The impact may be seen in Uygur and Kazak literacy, education, literary tradition, culture, and relationship with the majority Han. Regarding literacy, 33 percent of Uygurs and 39 percent Kazaks have some degree of difficulties in reading in the current Arabic script. For example, during research on this topic, the author contacted a few college Kazak language teachers, some of whom were not able to read fluently in the Arabic script. To make matters worse, older and younger Uygur and Kazak speakers also have great difficulty in reading works in the Latin scripts published between 1960 and 1982. These literacy problems especially affect education in Uygur and Kazak communities. One of the most serious problem in education in EMN regions is the lack of qualified teachers, who have the appropriate degree from college or, if without a college degree, can pass state qualifying examinations in general education and in the subjects that they have been teaching. The State Nationality Affairs Commission's documents show that, in many EMN regions, 40 to 50 percent, in some cases even 80 percent, of teachers are not qualified (see AOPROSNAC 1996). In Xinjiang, during 1984-1987, 50 percent of elementary school teachers, 79 percent of middle school teachers, and 77 percent of high school teachers were not qualified (Jia and Cai 1996). In Zhaosu County of Xinjiang where the Kazaks are the dominant nationality, for instance, only 83 out of

292 Kazak elementary school teachers, 8 out of Kazak 117 middle school teachers and 10 out of 45 Kazak high school teachers were qualified in the 1980s (Wang 1990). Though factors other than scripts also contribute to the quality of teachers, the low quality of teachers definitely results in low quality in teaching. In Kashi Prefecture's Yingjiesha County, out of 230 Uygur elementary school students, only· five passed the entrance examination to middle school (Wang 1990). The reverse may be seen in the majority Han communities, where the majority of elementary school students can pass the entrance examination to middle schools. The script reforms' direct impact on the quality of education may be seen in the lack of teacher reference books and student supplementary reading materials in Uygur and Kazak communities (Wang 1990:487). The situation becomes worse when the existing materials are in different scripts. For example, in Kazak, materials published before the early 1950s are in the old Arabic alphabet, those imported from the former Soviet Union during the middle of the 1950s are in the Cyrillic alphabet, those published during the 1960s and 1970s are in the Latin alphabet, and the most recent publications are in the revised Arabic script. Most of these publications seem like a foreign language to many teachers and all students.

The script reforms may also have a negative impact on cultural, intraethnic, and interethnic relationships. A third of the Uygur and Kazak population have trouble in reading publications in the Arabic scripts, a handicap that severs their ties with their ethnic heritage or at least erects some barriers between these generations and their ethnic heritage. This may not be the intention of the Chinese government, but the Kazaks might so perceive it, because quick assimilation into the Hans certainly was a government goal during the late 1950s, the 1960s, and the early 1970s. This perception has an adversarial impact on the Uygur and Kazaks' relationship with the majority Han, as was evidenced in 1962, when 100,000 to 200,000 Kazaks and Uygurs sought refuge in the former Soviet Union (cf. Benson and Svanberg). At least, it is now clear that the script reforms did not bring the Uygurs and Kazaks closer to the Hans, as the Chinese government intended, but rather alienated them. After so many years, this wound is still being healed (cf. Zhou in press a).

5. Conclusion

At a time of globalization, when ethnic minority languages and writing systems face critical challenges in their maintenance and survival, at least two lessons may be learned from the Chinese experience. First, the Chinese case seems to suggest that language policy should not be directly bundled with political and economic policies, though these policies will interact with each other on their own (cf. Zhou in press b). Simply put, in making political and economic policies, a government can afford to look into the future for only a few years, whereas in making language policy it has to look into the future for several decades or more. In China, the direct dictation of EMN language policy by the CCP's political and economic policies created negative impacts, while the relative independence of the language policy for Chinese (to promote PTH over dialects) appears to have had more positive results, perhaps because the latter has not been linked to any timetable in the CCP's agenda in the last forty-plus years (cf. Zhou 1999b, Zhou in press b).

Second, the Chinese case also suggests that great caution needs to be used and research has to be done on the individual· speech community's needs and support for script reform. Even if the Chinese EMN language policy is considered a failure in the cases of the Uygur and Kazak script reforms, the same policy may be considered more successful in the reforms of the Thai, Yi, and Miao writing systems and in the creation of writing systems for oral languages (Dai, Bi, and Wang 1990,

Shamajiajia 1989). These different results produced by the same EMN language policy are due more to the differences in linguistic and script situations than to the differences in the cultural, social, and political situations in those communities. But, a comparison of situations leading to the failure or success of the Chinese government's EMN language policy is beyond the scope of the current paper.

Note

References

Administration Office and Policy Research Office of the State Nationality Affairs Commission (AOPROSNAC). (1996). *Selected Documents of the State Nationality Affairs Commission* (Guojia minwei wenjian xianbian). Beijing: Aviation Press.

Azizi, Seypidin. (1960). Strive for the victory of socialist revolution in languages and scripts and for the unity and cooperation of all nationalities. *Xinjiang Education*, 4: 11-17.

Benson, Linda and Ingvar Svanberg. (1998). *China's Last Nomads: The History and Culture of China's Kazaks*. Armonk, NY: M. E. Sharpe.

Chomsky, Carol. (1991). Stages in language development and reading exposure. In *Language Issues in Literacy and Bilingual/multicultural Education*, ed. by Masahiko Minami and Bruce Kennedy, 48-72. Cambridge, MA: Harvard Educational Review.

Coulmas, Florian. (1989). *The Writing Systems of the World*. Oxford: Basil Blackwell.

Coulmas, Florian. (1996). *The Blackwell Encyclopedia of Writing Systems*. Oxford: Blackwell Publishers.

Dai, Qingxia, Xun Bi and Qingren Wang (eds.). (1990). *Language Relationships and Language Work* (Yuyan guanxi yu yuyan gongzuo). Tianjin: Classics Press.

Daniels, Peter and William Bright (eds.). (1996). *The World's Writing Systems*. Oxford: Oxford University Press.

Department of Population Statistics of State Statistical Bureau (DPSSSB) and Economic Department of State Nationalities Affairs Commission (EDSNAC) (1994) *Tabulation on China's Nationalities: Data on 1990 Population Census*. Beijing: China Statistical Publishing House.

Dreyer, June Teufel. (1976). *China's Forty Million: Minority Nationalities and National Integration in the People's Republic of China*. Cambridge, MA: Harvard University Press.

Dreyer, June Teufel. (1997). Assimilation and accommodation in China. In *Government Policies and Ethnic Relations in Asia and the Pacific*, ed. By Michael E. Brown and S3umit Ganguly, 351-391. Cambridge, MA: MIT Press.

Editorial Board of Contemporary Chinese Nationality Work (EBCCNW). (1989). *Chronicle of Major Nation/Nationality Issues in Contemporary China* (Dangdai zhongguo minzu gongzuo da shiji). Beijing: Ethnic Publishing House.

Geng, Shimin. (1989). *Modern Kazak Grammar* (Xiandai Hasakayu yufa). Beijing: Central University for Nationalities Press.

Hu, Zhenghua. (1979). Uygur writing systems. Ethnic Minority Languages (minzu yuwen), **2**: 152-157.

Jia, Chunzen and Qingsheng Cai. (1996). *Introduction to Ethnic Sociology* (Minzu shehuixue gailun). Beijing: Central University for Nationalities Press.

Johnson, Lars and Eva Agnes Csato2 (eds.) (1998). *The Turkic Languages*. London & New York: Routledge.

Liu, Xianzhao and Conggen Zhang. (1994). *Speeches on Nation/Nationality Issues by Chinese Communist Party Leaders* (Zhonggu gongchandang zhuyao lindaoren lun minzu wenti). Beijing: Ethnic Publishing House.

Mackerras, Colin. (1994). *China's Minorities: Integration and Modernization in the Twentieth Century.* Hong Kong: Oxford University Press.

Niu, Ruji. (1997). *An Introduction to Classical Uygur Scripts and Documents* (Weiwuer gu wenzi yu gu wenxian daolun). Urumqi: Xinjiang People's Press.

Script Office of the Commission on Writing System Reform of Xinjiang Uygur Autonomous Region (SOCWSRXUAR). (1962). On Uygur and Kazak writing system reform. *Xinjiang Education*, **1**: 35-36; **2**: 40-41.

Snow, Catherine. (1991). Literacy and language: The relationships during the preschool years. In *Language Issues in Literacy and Bilingual/ Multicultural Education*, ed. by Masahiko Minami and Bruce Kennedy, 207-232. Cambridge, MA: Harvard Educational Review.

Policy Research Office of the State Commission on Language and Script (PROSCLS). (1996). *Collection of Laws and Regulations on Language and Script: 1949-1995* (Guojia yuyan wenzi zhengce fagui huibian). Beijing: Language Press.

Seybolt, Peter J. and Gregory Kuei-ke Chaing (eds.). (1979). *Language Reform in China: Documents and Commentary.* New York: M.E. Sharpe, Inc.

Shamajiajia. (1989). *Selected Papers on EMN Language Teaching* (Shaoshu minzu yuwen jiaoxue lunwen ji). Huhhot: Inner Mongolia Education Press.

Stalin, Joseph. (1975). *Marxism and the National-Colonial Question.* San Francisco: Proletarian Publishers.

Xinjiang Lishi Jiaocai Bianxie Zu (XLJBZ, Editorial Group of Xinjiang History). (1993). *A Local History of Xinjiang* (Xinjiang difang shi). Urumqi: University of Xinjiang Press.

Wang, Xihong (ed.). (1990). Ethnic Minority Education along China's Borders (Zhongguo bianjing minzu jiaoyu). Beijing: Central University for Nationalities Press.

Zhou, Enlai. (1979). The immediate tasks in writing reform. In *Language Reform in China: Documents and Commentary*, ed. by Peter J. Seybolt and Gregory Kuei-ke Chaing, 228-43. New York: M.E. Sharpe, Inc.

Zhou, Enlai. (1994). Principles for reforms of writing systems for ethnic nationalities. In *Speeches on Nation/Nationality Issues by Chinese Communist Party Leaders*, ed. by Xianzhao Liu and Conggen Zhang, 190-191. Beijing: Ethnic Publishing House.

Zhou, Minglang. (1999a). Language policy and its impact on reforms of writing systems for minority languages in China. Manuscript. University of Colorado at Boulder.

Zhou, Minglang. (1999b). Attitudes of Cantonese and Xiang speakers toward Mandarin and its stereotyped speakers. *Texas Linguistics Forum*, **42**: 263-274.

Zhou, Minglang. (in press a). The official national language and language attitudes of three ethnic minority groups in China. *Language Problems and Language Planning*.

Zhou, Minglang. (in press b). Two developments of language spread under one language planning in China. In *Bilingualism: Bilingual Individuals and Bilingual communities*. University of Vigo Press.

BERKELEY LINGUISTICS SOCIETY

Book Prices*

Parasession & Special Session (S) Topics		Indiv.	Insitut.
BLS-1 (1975)		$10.00	$15.00
BLS-2 (1976)		$10.00	$15.00
BLS-3 (1977)		$10.00	$15.00
BLS-4 (1978)		$10.00	$15.00
BLS-5 (1979)		$10.00	$15.00
BLS-6 (1980)		$10.00	$15.00
BLS-7 (1981)		$10.00	$15.00
BLS-8 (1982)		$14.00	$20.00
BLS-9 (1983)		$10.00	$15.00
BLS-10 (1984)	Subordination	$14.00	$20.00
BLS-11 (1985)	Poetics, Metrics, and Prosody	$14.00	$20.00
BLS-12 (1986)	Semantic Typology	$14.00	$20.00
BLS-13 (1987)	Grammar and Cognition	$14.00	$20.00
BLS-14 (1988)	Grammaticalization	$14.00	$20.00
BLS-15 (1989)	Theoretical Issues in Language Reconstruction	$14.00	$20.00
BLS-16 (1990)	Legacy of Grice	$16.00	$20.00
BLS-16S (1990)	General Topics in American Indian Linguistics	$8.00	$12.00
Author Index to BLS 1-16S		$2.00 (w/ shipping)	
BLS-17 (1991)	The Grammar of Event Structure	$16.00	$20.00
BLS-17S (1991)	African Language Structures	$8.00	$12.00
BLS-18 (1992)	Place of Morphology in a Grammar	$16.00	$20.00
BLS-18S (1992)	Typology of Tone Languages	$8.00	$12.00
BLS-19 (1993)	Semantic Typology and Universals	$16.00	$20.00
BLS-19S (1993)	Syntactic Issues in Native American Languages.	$8.00	$12.00
BLS-20 (1994)	Dedicated to the Contributions of Charles J. Fillmore	$16.00	$20.00
BLS-20S (1994)	Historical Issues in African Linguistics	$8.00	$12.00
BLS-21 (1995)	Historical Issues in Sociolinguistics/ Social Issues in Historical Linguistics	$16.00	$20.00
BLS-21S (1995)	Discourse in Southeast Asian Languages	$8.00	$12.00
BLS-22 (1996)	The Role of Learnability in Grammatical Theory	$16.00	$20.00
BLS-22S (1996)	Historical Issues in Native American Languages	$4.00	$12.00
BLS-23 (1997)	Pragmatics and Grammatical Structure	$16.00	$20.00
BLS-23S (1997)	Syntax and Semantics in Africa	$8.00	$12.00
BLS-24 (1998)	Phonetics and Phonological Universals	$20.00	$25.00
BLS-24S(1998)	Indo-European Subgrouping and Internal Relations	$4.00	$8.00
BLS-25 (1999)	Loan Word Phenomena	$20.00	$25.00
BLS 25S (1999)	Caucasian, Dravidian and Turkic Linguistics	$8.00	$12.00