PROCEEDINGS OF THE THIRTY-SEVENTH ANNUAL MEETING OF THE

BERKELEY LINGUISTICS SOCIETY

February 12-13, 2011

SPECIAL SESSION

on

LANGUAGES OF THE CAUCASUS

Edited by

Chundra Cathcart
Shinae Kang
Clare S. Sandy

Berkeley Linguistics Society
Berkeley, CA, USA

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

ISSN 0363-2946

LCCN 76-640143

Published by eLanguage
http://elanguage.net/journals/bls

TABLE OF CONTENTS

A note regarding the contents of this volume

Karine Megerdoomian also presented a paper as part of the Special Session of the conference, though her work does not appear in this volume.

Foreword

We are pleased to present the proceedings of the BLS 37 Special Session, held at UC Berkeley in February 2011. We would like to thank the contributors to this volume and all those who attended and participated in the conference.

Clare S. Sandy

SPECIAL SESSION:
LANGUAGES OF THE CAUCASUS

A Comparative Phonetic Study of the Circassian Languages[1]

AYLA APPLEBAUM and MATTHEW GORDON
University of California, Santa Barbara

Introduction

This paper presents results of a phonetic study of Circassian languages. Three phonetic properties were targeted for investigation: voice-onset time for stop consonants, spectral properties of the coronal fricatives, and formant values for vowels.

Circassian is a branch of the Northwest Caucasian language family, which also includes Abhaz-Abaza and Ubykh. Circassian is divided into two dialectal subgroups: West Circassian (commonly known as Adyghe), and East Circassian (also known as Kabardian). The West Circassian subgroup includes Temirgoy, Abzekh, Hatkoy, Shapsugh, and Bzhedugh. East Circassian comprises Kabardian and Besleney. The Circassian languages are indigenous to the area between the Caspian and Black Seas but, since the Russian invasion of the Caucasus region in the middle of the 19th century, the majority of Circassians now live in diaspora communities, most prevalently in Turkey but also in smaller outposts throughout the Middle East and the United States.

1 Methodology

Results presented here are drawn from a total of 33 speakers. Of the 33, 26 hailed from Turkey, 4 from the Russian Federation, 1 from Syria, and 2 from Jordan. Nineteen consultants spoke Adyghe (18 from Turkey, 1 from Russia), 13 spoke Kabardian (7 from Turkey, 3 from Russia, 1 from Syria, 2 from Jordan), and 1 (from Turkey) spoke Besleney. The Adyghe speaker from Russia spoke the

[1] Thanks to the audience at the 37th meeting of BLS for feedback on this research. A special thanks to the many Circassian speakers without whose generosity and time this study would have been impossible. The research presented here was funded by an ELDP grant from the Hans Rausing Foundation to the first author and by NSF grant BCS0553771 to the second author.

literary (Temirgoy) variety, while the Adyghe consultants from Turkey self-reported as speaking the following dialects: Abzekh (4 speakers), Hatkoy (7 speakers), Shapsugh (5 speakers), Bzhedugh (2 speakers). Most of the recordings were made in Ankara, Turkey during a series of trips conducted between 2007-2010 with some additional recordings made in Orange County, California and, in the case of one Adyghe speaker, in Leipzig, Germany.

A corpus of 196 words designed to illustrate the principle phonetic contrasts of the targeted languages was elicited from the East Circassian (Kabardian and Besleney) speakers, while a corpus of 256 words was recorded from the West Circassian (Adyghe) speakers. The list included all the phonemic contrasts reconstructed for proto-Circassian (Kuipers 1963, 1975). Each word was repeated twice by each speaker after being prompted with the Turkish equivalent for speakers living in Turkey, the English equivalent for speakers living in Orange county, and Kabardian for the Adyghe speaker from the Adyghe Republic of Russian Federation. Targeted consonants appeared in word-initial, intervocalic and word-final contexts, while vowels appeared in stressed syllables of mono- or di-syllabic words. Data were recorded as .wav files at 44.1 kHz onto a solidstate recorder (either a Marantz PMD660 or an Edirol R09) via a Shure SM10 headworn unidirectional microphone. Digital recordings were transferred to computer in preparation for acoustic analysis using Praat (Boersma & Weenink 2010).

2 Results

2.1 Voice-onset-time (VOT)

Proto-Circassian is reconstructed as having a four way laryngeal contrast in the stop series between voiced, voiceless unaspirated, voiceless aspirated, and ejective (Kuipers 1963, 1975). Most varieties of modern Circassian, including the East Circassian languages Kabardian and Besleney and most varieties of Adyghe have neutralized the contrast between voiceless unaspirated and voiceless aspirated stops while preserving the original ejectives. Most Shapsugh dialects of Adyghe, however, are reported by Kuipers (1963, 1975) and Smeets (1984) to preserve a four-way laryngeal contrast. Most speakers of the Hatkoy dialect of Adyghe recorded by us also appear to maintain the original contrasts.

For the present study, voice-onset-time was measured in two contexts: word-initially and intervocalically. Word-medial exemplars appeared in disyllabic words between stressed /aː/ and unstressed /ɐ/. Virtually all word-initial tokens appeared in the first (stressed) syllable of disyllabic words before the vowel /aː/, although for certain speakers monosyllabic words had to be substituted. Places of articulation for which the data were best controlled for were measured: bilabials for the initial tokens and denti-alveolars for the medial ones. Voice-onset-time was measured from a waveform in conjunction with a time-aligned spectrogram.

Figure 1 contains bar graphs showing the mean VOT values (in seconds) averaged across speakers for the measured stops as produced by speakers of the two Circassian varieties, Shapsugh and Hatkoy, that maintain a four-way laryngeal contrast. Ejectives are included as well since voice-onset-time is potentially used as a cue to their identity. Note that the whiskers delimit the range of values one standard deviation from the mean.

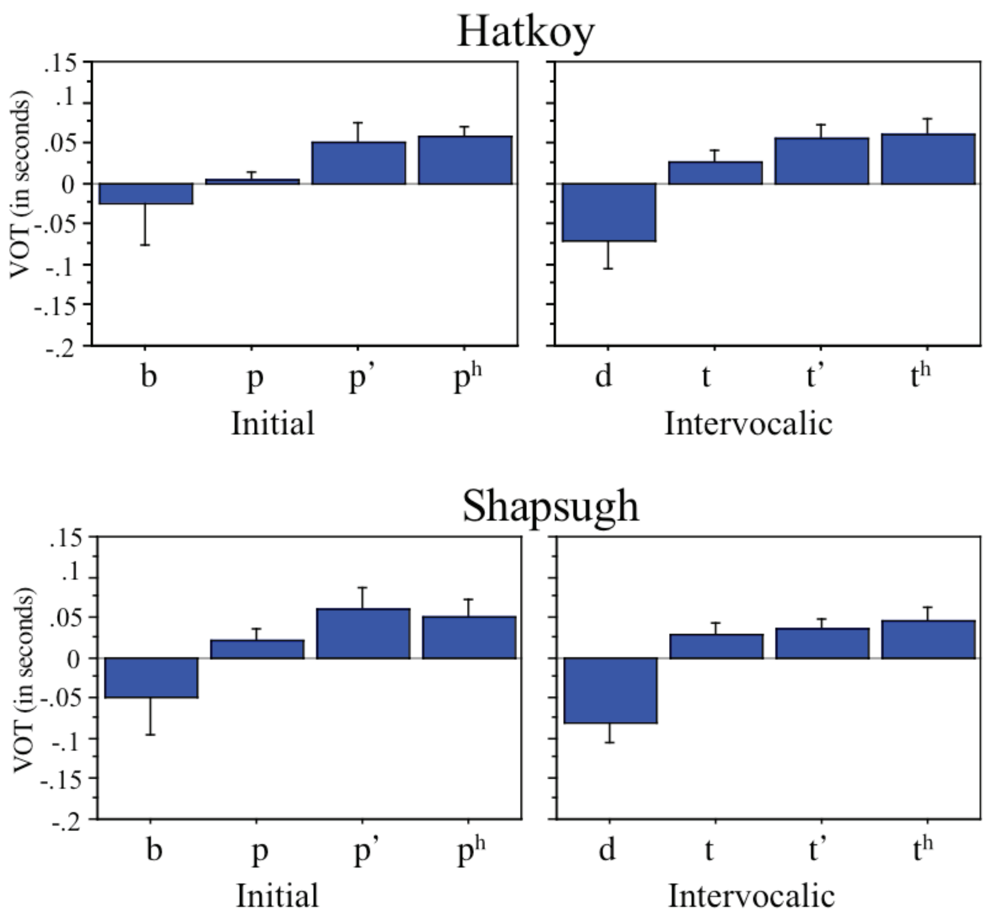


Figure 1. Voice-onset-time values (in seconds) for voiced, unaspirated, ejective, and aspirated stops in two contexts averaged over 7 Hatkoy and 5 Shapsugh speakers.

As figure 1 shows, the contrast between voiced, voiceless unaspirated and voiceless aspirated stops is preserved in both Hatkoy and Shapsugh, though the difference in VOT between the unaspirated and aspirated stops in both varieties is considerably smaller in intervocalic position, where the unaspirated stops have longer VOT values than in initial position.

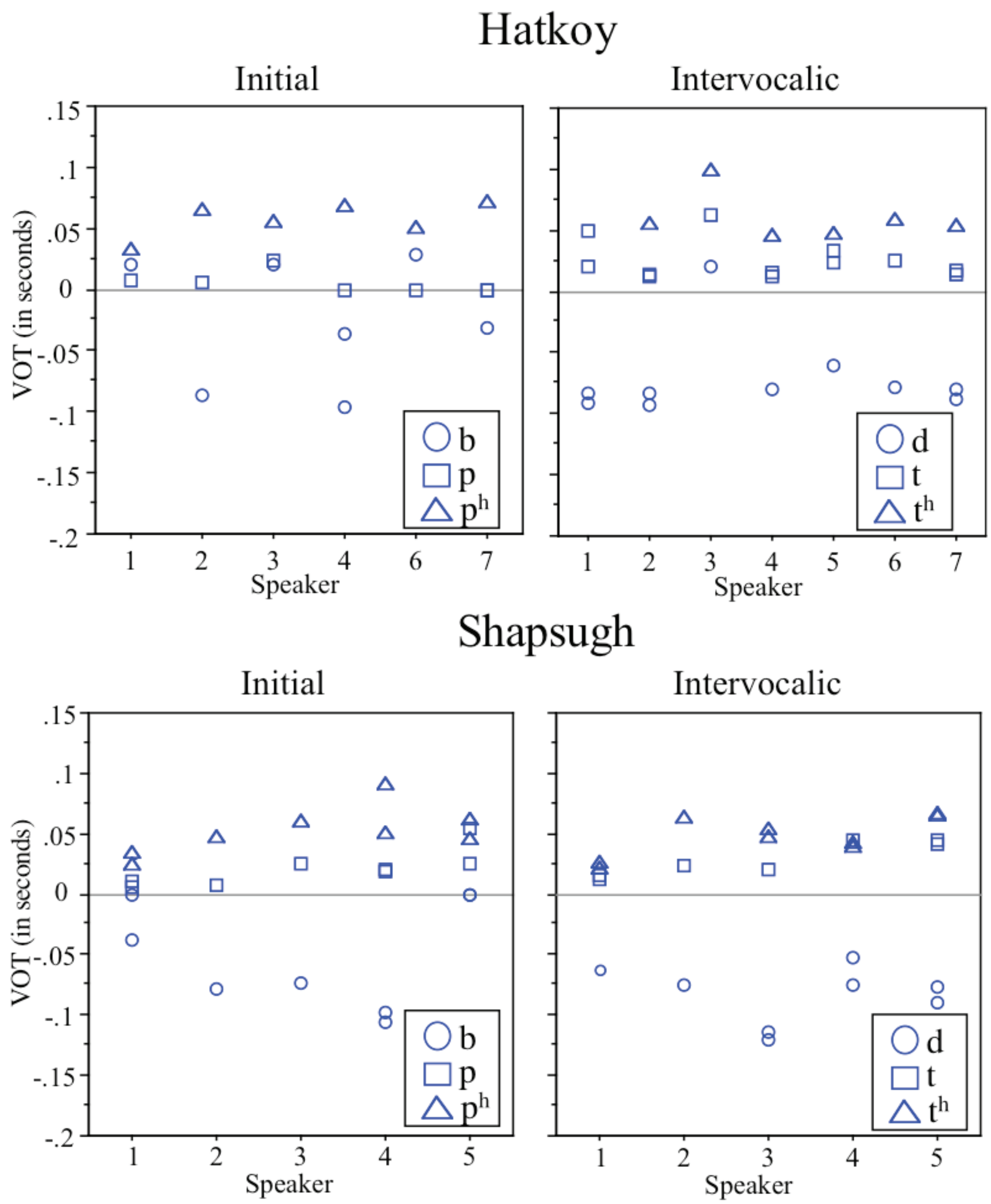


Figure 2. VOT values (in seconds) for tokens of voiced, unaspirated, and aspirated stops in two contexts for 7 Hatkoy and 5 Shapsugh speakers.

The situation is actually more complex than the across speaker means in figure 1 suggest, as certain speakers in both Hatkoy and Shapsugh appear to be collapsing at least two of the three non-ejective series. Figure 2 plots VOT for separate tokens of the voiced, unaspirated, and aspirated stops as produced by individual speakers of Hatkoy and Shapsugh. The three-way contrast is particularly vulnerable in initial position, where three of the seven Hatkoy speakers and two of the Shapsugh speakers produce the phonemic voiced stops without consistent prevoicing during the closure. This mode of realization infringes on the VOT space of the voiceless stops resulting in neutralization or near-neutralization word-initially. It is, in fact, unclear whether Hatkoy speaker 1 or Shapsugh speaker 1 contrast any of the non-ejective stops word-initially. (Note that Hatkoy speaker 1 did not produce the target word containing aspirated stops word-medially.) Word-medially Hatkoy speaker 5 and Shapsugh speakers 1, 4, and possibly 5 appear to have lost the VOT difference between aspirated and unaspirated stops.

Figure 3 confirms that the Bzhedugh and Abzekh varieties of Adyghe have neutralized the distinction between unaspirated and aspirated stops both initially and medially. Neutralization results in a stop with virtually no aspiration word-initially in Bzhedugh and one with some aspiration word-medially in Bzhedugh and both initially and medially in Abzekh.

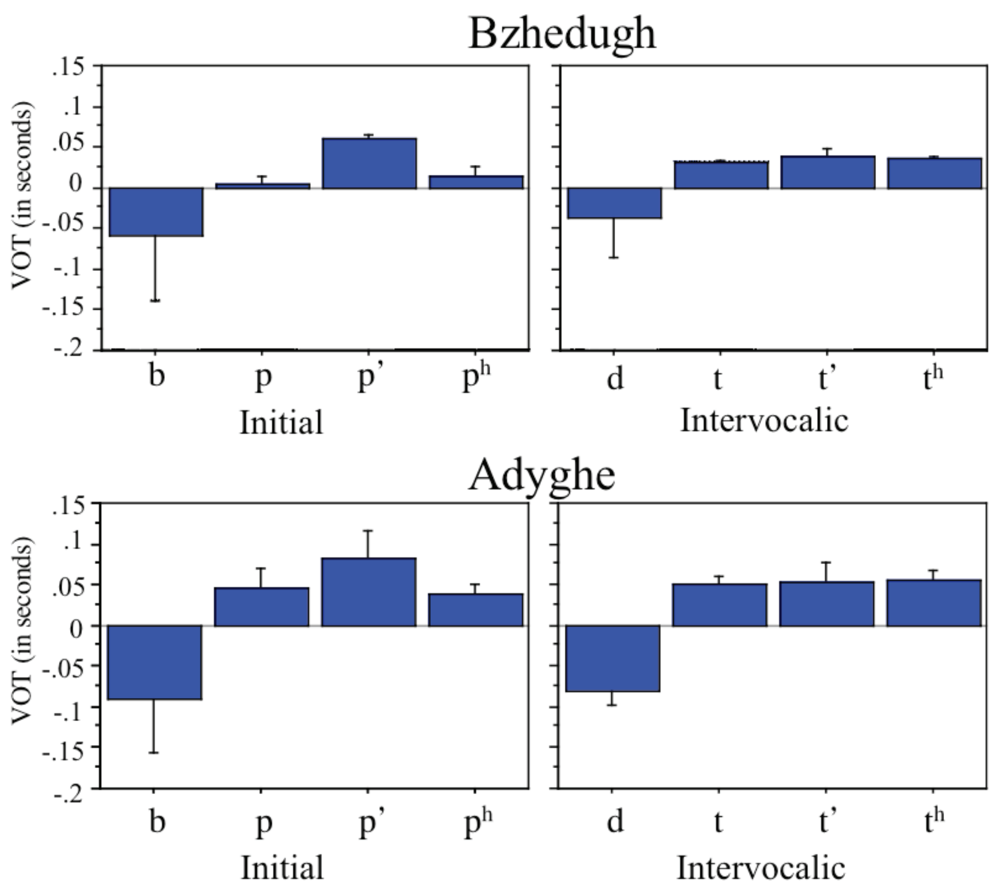


Figure 3. VOT values (in seconds) for voiced, unaspirated, ejective, and aspirated stops averaged over 2 Bzhedugh and 4 Abzekh speakers.

The proto-Circassian four-way contrast has also been lost in East Circassian (Kabardian and Besleney), which has collapsed the etymologically unaspirated and aspirated series word-initially (e.g. *paːnɐ 'thorn' and *pʰaːsɐ 'early' both begin with /p/ in East Circassian) and the etymologically unaspirated and voiced series intervocalically (e.g. *saːpɐ 'dust' and *xʷaːbɐ 'hot' both have /b/ in East Circassian). In both positions, as figure 4 shows, the contrast has been neutralized in favor of aspiration, plausibly to enhance the contrast with the phonemic voiced series. (Note that there is only word-initial stop data from Besleney and only from a single speaker.)

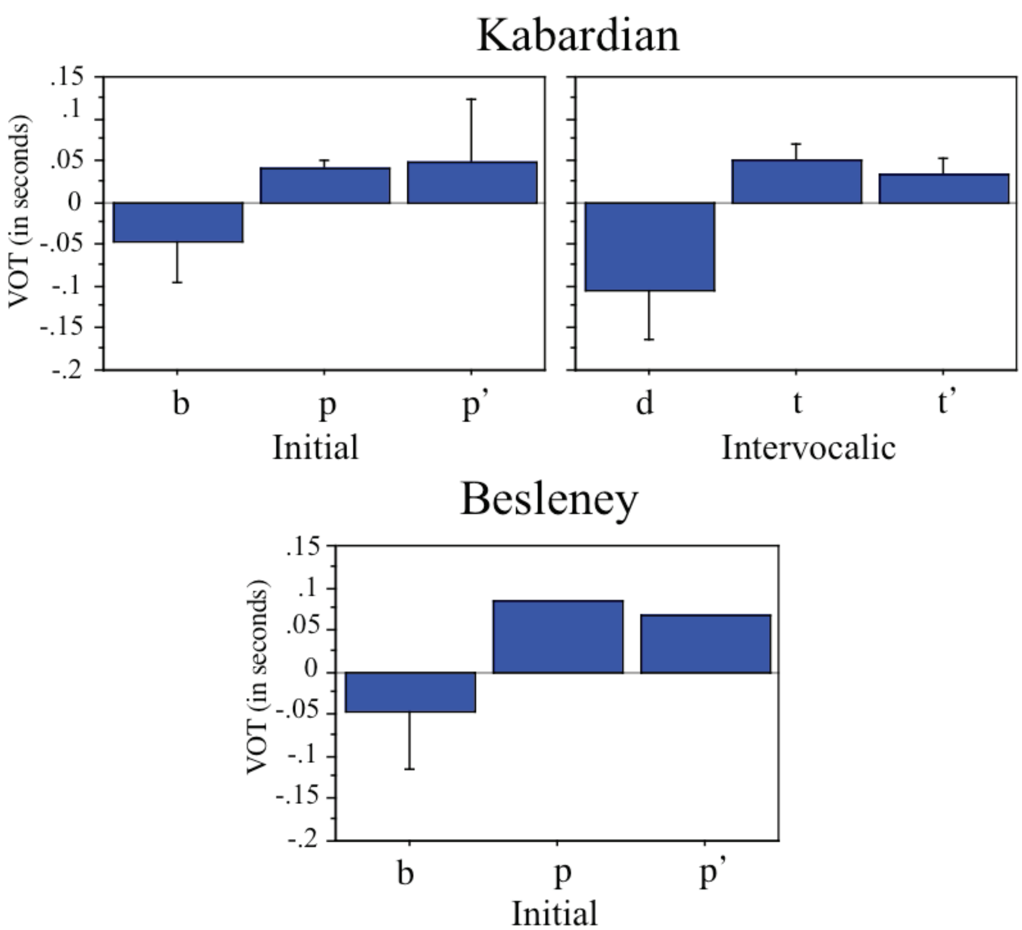


Figure 4. Voice-onset-time values (in seconds) for voiced, voiceless, and ejective stops in two contexts (one for Besleney) averaged over 13 Kabardian and one Besleney speaker.

2.2 Spectral Properties of Fricatives

Proto-Circassian is reconstructed by Kuipers (1963, 1975) as having 14 coronal fricatives in addition to fricatives at the bilabial, velar, uvular, pharyngeal and glottal places of articulation. Four laryngeal settings are reconstructed for the coronals by Kuipers: voiceless unaspirated, voiceless aspirated, ejective, and voiced. In addition, four places of articulation contrasted in the proto-language and in certain modern West Circassian varieties: denti-alveolar, alveolopalatal, and two postalveolar series. There is disagreement about the phonetic nature of the contrast between the two postalveolars. Kuipers (1963, 1975) describes it as a contrast between plain palatals and palatalized palatals, whereas Smeets (1984) characterizes it as a contrast between plain (Kuipers' palatalized palatals) and velarized (Kuipers' plain palatals). The latter series, i.e. Kuipers' plain palatals and Smeets' velarized palatals, is characteristically, though not exclusively, realized, as far as we can tell from acoustic data and observations about the articulation, as a domed postalveolar (palato-alveolar) fricative, i.e. /ʃ/, whereas the second series, i.e. Kuipers' palatalized palatals and Smeets plain palatals, varies in its realization across individuals and varieties. The most typical realization seems to be as a laminal closed postalveolar as described for the related Northwest Caucasian language Ubykh by Ladefoged & Maddieson (1996). Catford (ms cited in Ladefoged & Maddieson 1996:161) describes its production as follows: "acoustically and physiologically between a typical s and a typical ʃ:In its production the tip of the tongue rests against:the lower teeth (as for a laminal s̻), but the main articulatory channel is at the back of the alveolar ridge (as for a lamino-post-alveolar ʃ)" Catford (and Ladefoged & Maddieson 1996) transcribe it as ŝ, a transcription which we adopt here, and refer to it as a "closed laminal postalveolar fricative". It is "laminal" because the contact between the tongue and the upper surface of the mouth is relatively broad in the front-back dimension and "closed" because it is not produced with the sublingual cavity that often characterizes postalveolar fricatives cross-linguistically. The 14 coronal fricatives reconstructed for proto-Circassian by Kuipers (1963, 1975) are thus shown in table 1, using the transcriptions of /ʃ/ and /ŝ/ for the two postalveolars.

Table 1. The 14 coronal fricatives of proto-Circassian (Kuipers 1975) transcribed to approximate articulatory characteristics of their modern reflexes

	Dental	Alveolo-palatal	Post-alveolar[1]	Post-alveolar[2]
Voiceless	s	ɕ ɕʷ	ʃ	ŝ
Voiceless aspirated			ʃʰ	ŝʰ
Ejective		ɕ' ɕʷ'		
Voiced	z	ʑ ʑʷ	ʒ	ẑ

In our data, all 14 coronal fricatives are found only for one of the Shapsugh speakers and not for any speakers of the other Circassian varieties. Elsewhere, a subset of contrasts is found with the phonetic nature of this contrast and the number of contrasts varying from variety to variety and often from speaker to speaker.

One of the typologically rare features of the Circassian fricative inventory is the aspirated postalveolar fricative, which is synchronically preserved only for certain Shapsugh speakers in our data.[2] Even for those speakers maintaining a contrast between aspirated and unaspirated fricatives, aspiration is limited to certain lexical items and has been lost in many words that are etymologically expected to contain an aspirated fricative. The word for 'horse', /ʃʰə/, is the lexical item most reliably associated with aspiration, perhaps because it is a high-frequency word that differs minimally through its aspiration from another high-frequency lexeme, the word for 'three', /ʃə/. Aspiration is often associated with nasalization as Colarusso (1988) observes for Bzhedugh (although our Bzhedugh speakers do not have aspirated fricatives). Figure 5 shows the contrast between an aspirated fricative in /ʃʰə/ 'horse' and an unaspirated fricative in /ʃədə/ 'donkey' as produced by a female Shapsugh speaker.

Languages vary in how many of the 14 coronal fricatives from proto-Circassian are preserved synchronically. If we take the voiceless series as representative, all languages have /s/ but varieties differ in whether they have one, two, or three additional coronal places of articulation represented. A common theme is for Diaspora speakers outside of Russia to have fewer contrasts. At the extreme end of simplicity, Besleney and Turkish Kabardian (and the Baksan Russian Kabardian speaker) neutralize /ɕ/, /ʃ/ and /ŝ/ to /ʃ/. Russian Kabardian (and the Kabardian speaker from Syria), Hatkoy, Bzhedugh, and Diaspora Adyghe occupy middle ground possessing two coronal places posterior to /s/. These two places vary depending on speaker: /ɕ/ vs. /ʃ/ or /ŝ/ vs. /ʃ/ or /ʂ/ vs. /ʃ/. At the extreme end of complexity, Shapsugh and Temirgoy Adyghe preserve a 3-way posterior coronal contrast, where the phonetic nature of the contrast shows considerable interspeaker and interdialectal variation. Figure 6 shows a spectrogram (top) and FFT spectra (bottom) of fricatives involved in the four-way contrast between /s/, /ʂʰ/, /ʃ/ and /ŝ/ as produced by a speaker of Shapsugh from Turkey. For this speaker of Shapsugh, the four coronals are distinguished relatively clearly through the distribution of their noise in the frequency domain. The denti-alveolar /s/ is associated with the highest frequency energy, much of which is above the 8000Hz upper limit of the spectrogram. Proceeding from the retroflex fricative /ʂʰ/ in 'horse' to the domed postalveolar /ʃ/ in 'milk' to the laminal closed postalveolar /ŝ/ in 'hundred', the lower limit of the primary locus

---

2 Colarusso (1988) mentions a uvular aspirated fricative in Bzhedugh found only in the plural morpheme /χʰa/ but our Bzhedugh speakers have a simple unaspirated uvular fricative in this morpheme.

of energy progressively increases in frequency. Note that the word for ‘hundred’ begins with a sound that is phonetically quite similar to the sound in ‘horse’ for some speakers.

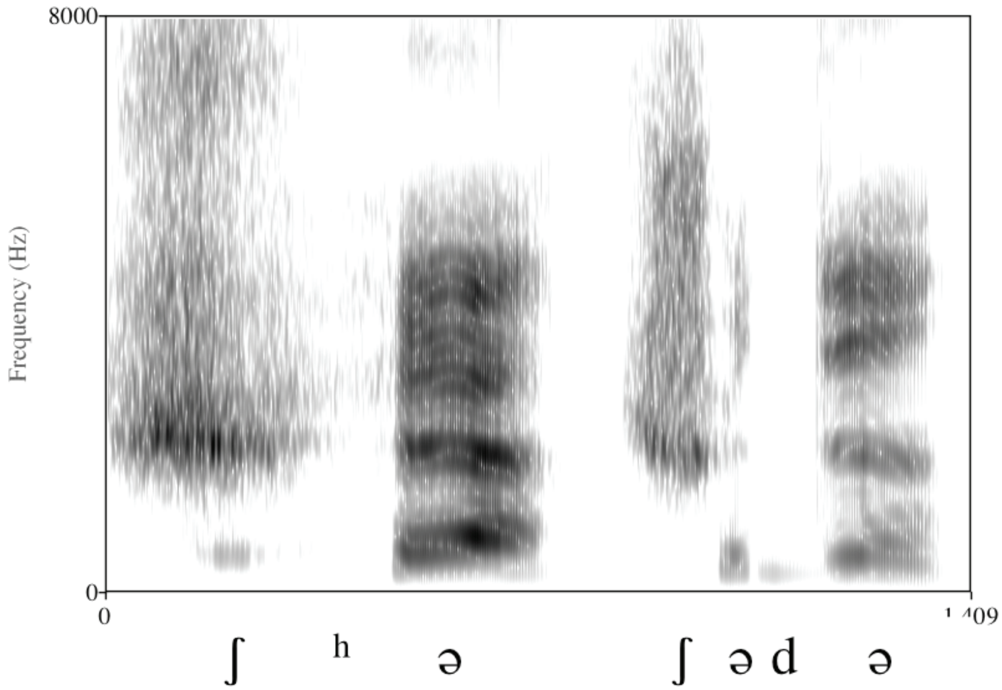


Figure 5. The aspirated fricative in /ʃʰə/ ‘horse’ vs. the unaspirated fricative in /ʃədə/ ‘donkey’ as produced by a female Shapsugh speaker.

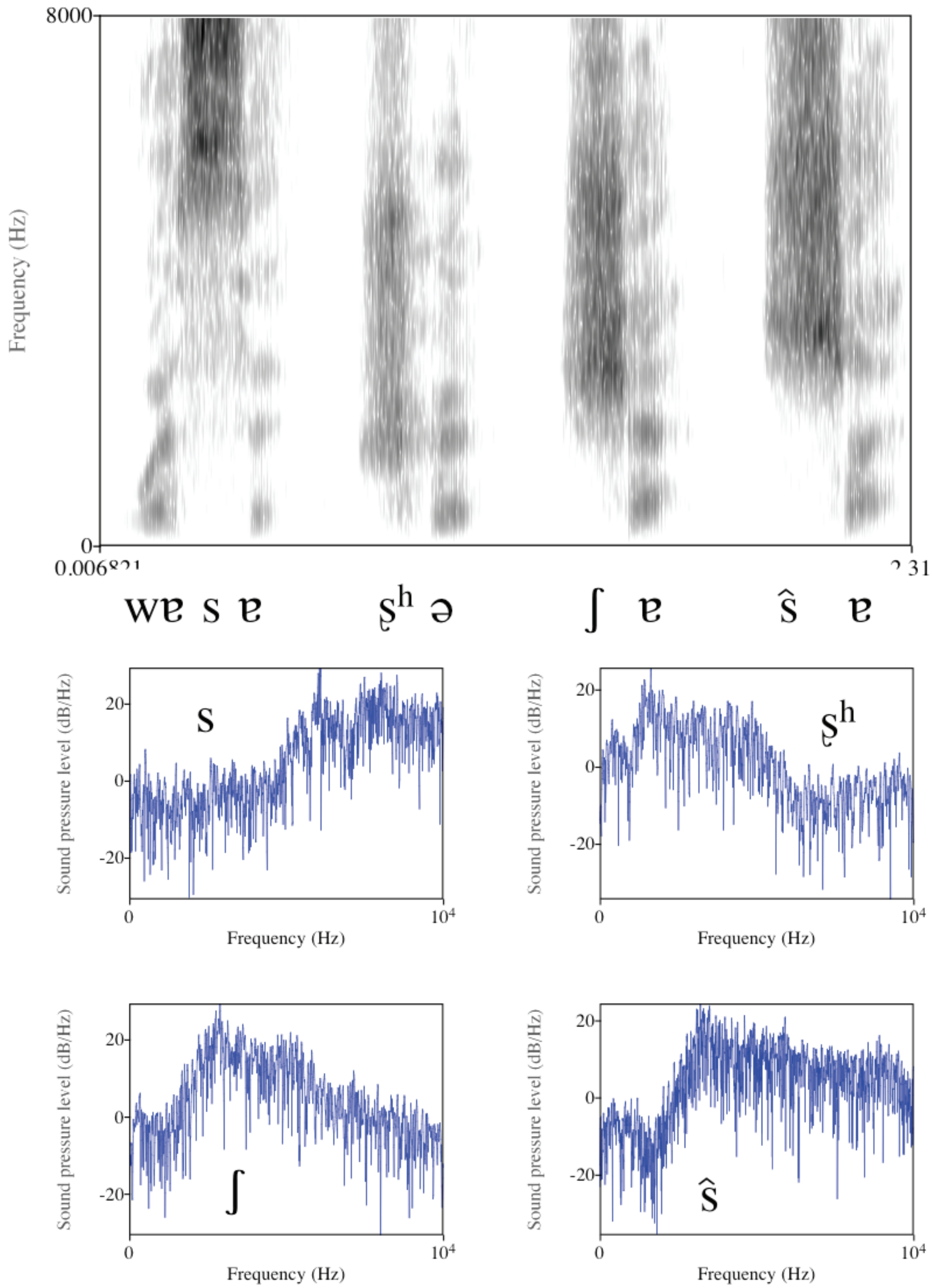


Figure 6. Spectrogram (top) and FFT spectra (bottom) of the four coronal fricatives in the words /wɐsɐ/ ‘snow’, /ʂ^{h}ə/ ‘horse’, /ʃɐ/ ‘milk’, and /ŝɐ/ ‘hundred’ uttered by a Shapsugh speaker.

Figure 7 shows representative FFT spectra from an Adyghe speaker who also distinguishes four coronal places of articulation. A similar increase in energy going from retroflex to domed postalveolar to closed postalveolar is observed here as well.

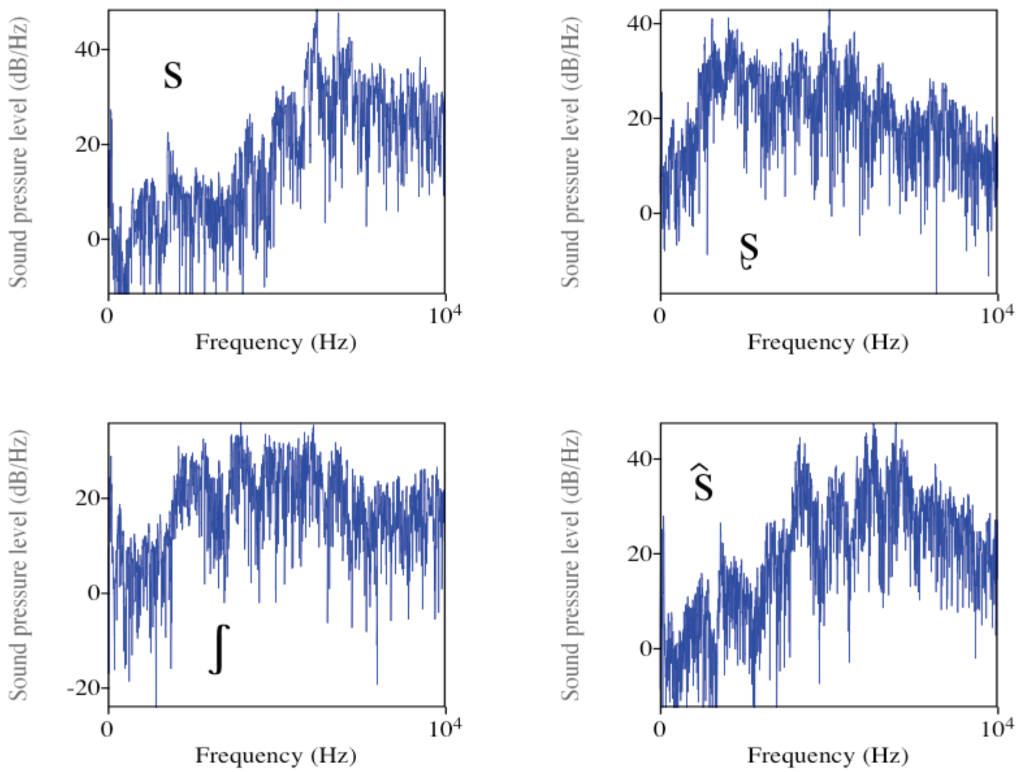


Figure 7. FFT spectra for the four coronal fricatives in the words /pʰaːsʁ/ ‘early’, /ʂə/ ‘horse’, /gʷaːʃʁ/ ‘princess’, and /məŝʁ/ ‘bear’ uttered by a Temirgoy Adyghe speaker.

To summarize the fricatives, Circassian varieties range from preserving all the relatively subtle coronal contrasts to having only a two-way distinction between denti-alveolars and postalveolars. There is considerable variation in the direction of neutralization. The fricative (the one in ‘horse’) labeled (depending on the source) either as a velarized palatal or a plain non-palatalized palatal is particularly prone to neutralizing with another series, though the direction of this merger varies. It may collapse with the alveolopalatal (e.g. in ‘hundred’) or with the domed postalveolar (e.g. in ‘milk’). The direction of the merger is likely attributed to variation in its production. The retroflex realization is acoustically similar to the domed postalveolar, whereas the closed laminal postalveolar realization is more like the alveolopalatal. In any case, the close proximity of the alveolopalatal, the domed postalveolar, and the third postalveolar (the velarized/non-palatalized one in 'horse') makes the three-way contrast unstable and prone to neutralization. Even the two-way contrast between alveolopalatal and postalveolar is neutralized in Turkish Kabardian.

2.3 Vowels

Most analyses of Circassian languages (e.g. Yakovlev 1948, Turchaninov & Tsagov 1940, Apazhev et al. 1957, Abitov et al. 1957, Bagov et al. 1970) assume three underlying vowels (2 short and one long) and four additional surface long vowels that are underlyingly short vowel + glide sequences, as shown in table 2.

Table 2. A representative vowel inventory for Circassian languages.

	Front	Central	Back
High	iː /əj/		uː /əw/
Mid	eː /ɐj/	ə ɐ	oː /ɐw/
Low		aː	

There are very few differences in vowel quality that can reliably be attributed to dialect as opposed to idiolect. One of the interesting issues is the spacing of the vowels in the height dimension since Circassian languages have "vertical" vowel systems. Figure 8 plots the three phonemic central vowels in the height (first formant) and backness/frontness (second formant) dimensions for the Turkish Kabardian speakers, for whom the data set is largest and whose vowel spacing is representative of other Circassian varieties.

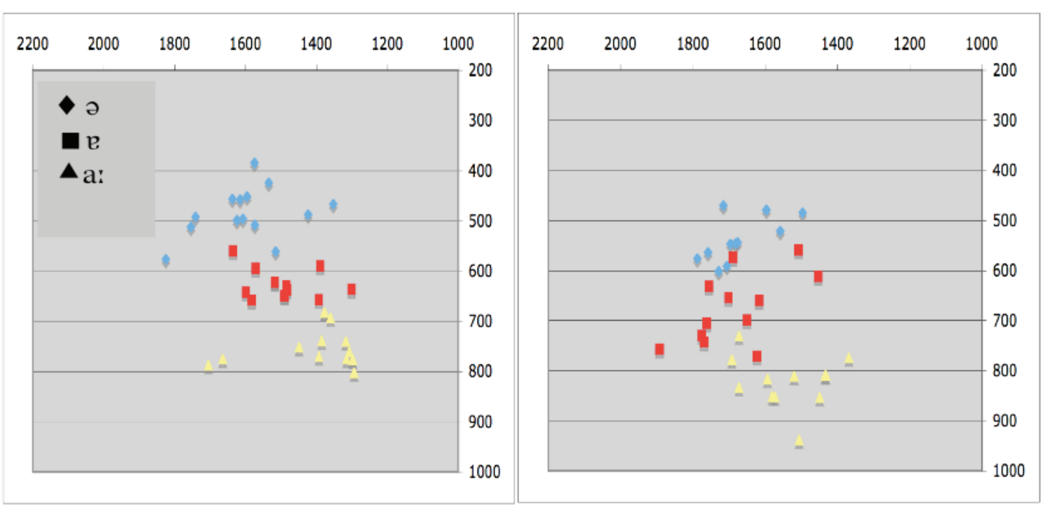


Figure 8. First vs. second formant plots (in Hertz) for male (left) and female (right) speakers of Turkish Kabardian.

The three vowels are fairly well differentiated in the vertical dimension and are consistent with the transcription of them as /ə/, /ɐ/, and /aː/. Figure 9 plots the vowel space for four male Adyghe speakers (three from Turkey and one from Russia).

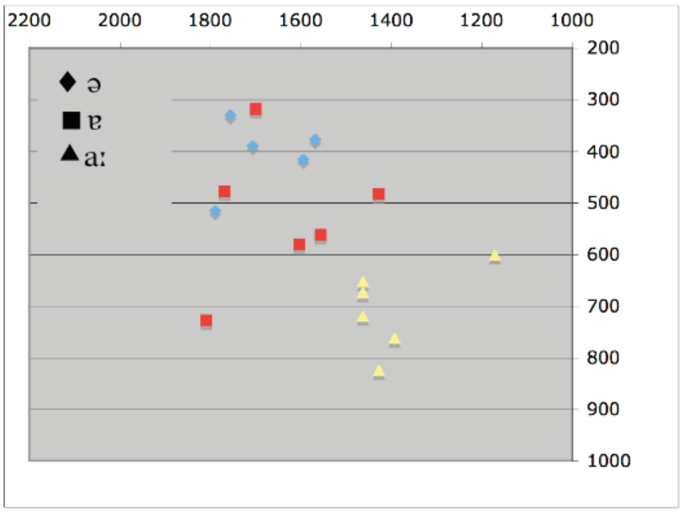


Figure 9. First vs. second formant plots (in Hertz) for four male speakers of Adyghe.

The two higher central vowels are shifted slightly upward in height relative to the male Turkish Kabardian data suggesting a transcription of these vowels as /ɨ/ and /ə/. The lowest vowel is also slightly retracted relative to the other two in figure 10. Note also the outlier data point for the mid vowel in the middle of the highest vowel's space.

It is interesting to note that Catford (1984) and Choi's (1991) studies of Kabardian show first formant values for the two higher central vowels in the Terek variety of Russian Kabardian that are similar to those in our Adyghe data and lower than those found in our Turkish Kabardian data. On the other hand, Wood's (1994) study of vowels in the Kuban dialect of Russian Kabardian produced results that are compatible with our Turkish Kabardian results.

3 Conclusions

Circassian languages are typologically unusual in the complexity of the fricative inventories, particularly in the coronal subspace. This phonetic complexity, however, has lead to instability in the realization of the contrasts. Many dialects and idiolects collapse certain of the original contrasts with the direction of neutralization varying considerably. A similar phonetic complexity is observed in the laryngeal contrasts of proto-Circassian with a similar result synchronically:

instability and neutralization. The neutralization of phonetically subtle contrasts has likely been further aided by the gradual erosion of the native speaker populations of these languages, many of which are seriously endangered, in the face of pervasive contact with other socially, economically, and politically dominant languages. Contact between different Circassian dialects has also likely led to cross-dialect influence on the production of certain phonemic contrasts. The vowel systems, which stand in sharp contrast to the fricative and stop inventories in their phonetic simplicity, have been preserved throughout the Circassian family.

References

Abitov, M. L., B. X. Balkarov, J. D. Desheriev, G. B. Rogava, X. U. El'berdov, B. M. Kardanov, and T. X. Kuasheva. 1957. *Grammatika Kabardino-Cherkesskogo literaturnogo jazyka*. Moscow: Izdatel'stvo Akademii Nauk.

Apazhev, M. L., N. A. Bagov, P. M. Bagov, B. X. Balkarov, J. N. Kokov, X. X. Zhakamyxov, and X. Sh. Urusov. 1975. *Kabardinsko-Russkij Slovar'*. Moscow: Gosydarstvennoe Izdatel'stvo Inostrannyx I Natsional'nyx Slovarej.

Bagov, P. M., B. X. Balkarov, T. X. Kuasheva, M. A. Kumaxov, and G. B. Rogova, eds. 1970. *Grammatika Kabardino-Cherkesskogo Literaturnogo Jazyka: 1 Fonetika i Morphologija*. Moscow: Nauka.

Boersma, Paul and David Weenink. 2010. *Praat: Doing phonetics by computer* (version 5.1.42) (available at www.praat.org).

Catford, J. C. 1984. Instrumental data and linguistic phonetics. In Jo-Ann W. Higgs and Robin Thelwall, eds., *Topics in Linguistic Phonetics*, *In Honour of E. T. Uldall*, 23-48, Coleraine, N. Ireland: New University of Ulster.

Choi, John. 1991. An acoustic study of Kabardian vowels. *Journal of the International Phonetic Association* 21:4-12.

Colarusso, John. 1988. *Northwest Caucasian Languages: A Phonological Survey*. New York: Garland.

Kuipers, A. H. 1963. Proto-Circassian phonology: An essay in reconstruction. *Studia Caucasica* 1:56-92.

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

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

Smeets, Riks. 1984. *Studies in West Circassian Phonology and Morphology*. Leiden: The Hakuchi Press.

Turchaninov, G. and M. Tsagov. 1940. *Grammatika Kabardinoskogo Jazyka*. Moscow: Izdatel'stvo Akademii Nauk.

Wood, Sidney. 1994. A spectrographic analysis of vowel allophones in Kabardian. *Working Papers* 42, 241—250. Lund: Lund University Department of Linguistics.

Yakovlev, N. F. 1948. *Grammatika Literaturnogo Kabardino-Cherkesskogo Jazyka*. Moscow: Izdatel'stvo Akademii Nauk.

Ayla Applebaum
Matthew Gordon
Department of Linguistics
University of California,
Santa Barbara, CA 93106

ayla1@umail.ucsb.edu
mgordon@linguistics.ucsb.edu

Cyclic Agreement and Empty Slots in Pazar Laz*

ÜMİT ATLAMAZ
Boğaziçi University

Introduction

This paper discusses the compatibility of templatic morphology and cyclic agreement on verbal agreement prefixes in Pazar Laz. It is based on templatic morphology and introduces the following questions: *Can agreement slots on verbal agreement remain empty through the steps of derivation? Is there insertion of a dummy element in cases when arguments are deficient in terms of agreement?* The organization of the paper is as the following: It first introduces the relevant background information about templatic morphology and then, it presents data from Pazar Laz to show that it has a templatic morphology on verbs. In section 3 we propose a cyclic agreement model based on Bejar & Rezac (2009) and discuss with relevant data. Section 4 summarizes and concludes the paper.

1 Templatic Morphology & Pazar Laz

1.1 Background

Languages displaying an invariant order of morphemes on a word and mutual exclusivity of morphemes despite semantic compatibility have been claimed to have a templatic morphology (Stump 2006, Inkelas 1993). According to Spencer (1991) there are certain languages, where a word consists of a stem and some other obligatory affixes that go into certain slots defined by a template. One rather crucial work on templatic morphology which will help us determine whether Pazar Laz has a templatic morphology or not comes from Inkelas (1993), where she shows complementary appearance of certain morphemes despite semantic

* I would like to express my gratitude to Balkız Öztürk for her invaluable feedback and support throughout my experience with Pazar Laz, 37th annual meeting of BLS, and the current paper. It is also my duty to thank my Pazar informant İsmail Avcı Bucak'lişi for the Laz data he generously provided me with.

compatibility. As a result, she argues that certain slots are reserved for certain morphemes. Once a slot is filled with a morpheme, all the other candidates are blocked.

In light of Inkelas (1993), Stump (2006), and Spencer (1991), we observe that Pazar Laz verbs display properties of templatic morphology. Certain morphemes go into certain slots. Table (1) is a simplified version of a verb template in Pazar Laz, which is enough to serve the purpose of the current paper.[1]

(1)

Prefixes			**Root**	**Suffixes**			
I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**
PV	*1. IO Agr.* *2. DO Agr.* *3. Subj. Agr.*	1. Reflexive 2. Causative 3. High Applicative 4. Low Applicative	Root	Series Marker	Modal	Subj. Agr.	Number Agreement

Slot II in Table (1) houses a competition among the person agreement markers. These morphemes are in complementary distribution despite their semantic compatibility. As a result of this complementary distribution and the relative places of other morphemes on the verb we argue that Pazar Laz verbs show templatic morphology. Sentences (2), (3), and (4) show that slot II cannot be filled by more than one marker.

(2) Ma **v**-inçir-i
I.ERG 1Subj-swim-1sgPAST
'I swam.'

(3) Ma si ce-**k**-ç-i.
I.ERG you.NOM PV-2Obj-beat-1sgPAST
'I beat you.'

(4) *Ma si ce-**p**-**k**-ç-i
I.ERG you.NOM PV-1Subj-2Obj-beat-1sgPAST
'I beat you.'

Sentence (4) indicates that there is enough space for only one agreement marker in slot II in Pazar Laz despite the compatibility of 1Subj and 2Obj in terms of

[1] See Öztürk & Pöchtrager (2011) for the full represenation of slots that can appear on the verbal complex in Pazar Laz.

meaning. Therefore we conclude that Pazar Laz shows templatic morphology for verbs.

2 Verbal Agreement in Pazar Laz

Before moving onto the theoretical discussion of cyclic agreement in Pazar Laz we would like to present some facts about Pazar Laz verbal agreement and some relevant data to be discussed in further sections.

Pazar Laz verbs carry person and number agreement, which are checked separately. Slot VII is reserved for structural subject (person only). Slot II houses a competition among arguments in terms of person agreement. Slot VIII houses number agreement. Sentences (5), (6) and (7) illustrate the agreement markers and their relevant places on a verb.

(5) Ma si ce-**k**-ç-**i**
I.ERG you.NOM(sg) PV-**2Obj**-beat -**1sgPAST**
'I beat you (sg).'

(6) Ma t'k'va ce-**k**-ç-**i-t**.
I.ERG you .NOM(pl) PV-**2Obj**-beat -**1sgPAST-pl**
'I beat you (pl).'

(7) T'k'va ma ce-**m**-ç-**i-t**
You.ERG (pl) me.NOM PV-**1Obj**-beat -**2sgPAST-pl**
'You (pl) beat me.'

Given that person and number are encoded separately, in the absence of overt pronouns, we observe ambiguity.[2]

(8) ce-**k**-ç-**i-t**.
PV-**2 Obj**-beat -**1sgPAST-pl**
a. 'I beat you (pl).'
b. 'We beat you (sg).'
c. 'We beat you (pl).'

Since the main focus of the current paper is on the verbal agreement prefix, the following data will concentrate on slot II.

[2] It should be noted that Pazar Laz is a pro-drop language and can drop both subjects and objects.

2.1 Slot II Competition

As discussed before, slot II can house different morphemes depending on the argument structure of the verb and deficiencies in arguments. Morphemes that can fill slot II are listed on table (9). The subsequent sections will provide sentences with different arguments and agreement patterns to display the competition for slot II.

(9)

	Intransitive Subject[3] (Agent/Theme)-Transitive Subject (Agent)	**Transitive Subject (Experiencer)**	**DO**	**IO**
1sg	v/f/p/b	m	m	m
2sg	Ø	k/g	k/g	k/g
3sg	Ø	Ø	Ø	Ø
1pl	v/f/p/b	m	m	m
2pl	Ø	k/g	k/g	k/g
3pl	Ø	Ø	Ø	Ø

Those morphemes listed under the second column on table (2) are usually referred to as ‘v-set’ whereas the rest are called as ‘m-set’ (Holisky 1991).

2.2 Intransitive Verbs

In sentences with intransitive verbs, slot II is filled with a ‘v-set’ marker unless the argument is deficient. Deficient, in this case, means that the argument does not have an overt agreement marker to fill slot II.

(10) **v**-inçir-i
1Subj-swim -1sgPAST
‘I swam.’

(11) **Ø**-inçir-i
Ø- swim -2sgPAST
‘You swam.’

(12) **Ø**-inçir-u
Ø- swim -3sgPAST
‘He/She swam.’

[3] See Öztürk 2008, 2010, Emgin 2009 and Öztürk & Pöchtrager (2011) for the general case and agreement patterns of subjects and objects in relation to their theta-roles in Pazar Laz.

2.3 Transitive Verbs with Agentive Subjects | Competition Starts

The competition for slot II surfaces in transitive sentences. Unless the object is deficient (i.e. third person), Slot II is filled with relevant DO maker. Otherwise, the subject fills Slot II with the relevant v-set agreement marker.

(13)	Si	ma	ce-**m**-ç-i.
	You.ERG	me.NOM	PV-**1Obj**-beat -2sgPAST
	'You beat me.'		

(14)	Ma	si	ce-**k**-ç-i.
	I.ERG	you.NOM	PV-**2Obj**-beat -1sgPAST
	'I beat you.'		
(15)	Ma	him	ce-**p**-ç-i.
	I.ERG	him/her.NOM	PV-**1Subj**-beat -1sgPAST
	'I beat him.'		

(16)	Si	him	ce-**Ø**- ç-i
	You.ERG	him/her.NOM	PV-**Ø**-beat -2sgPAST
	'You beat him/her.'		

2.4 Ditransitive Verbs

In sentences with ditransitive verbs, the privilege to fill slot II belongs to IO. When IO is deficient, the chance moves to DO and if that is deficient too, then ultimately the subject wins the competition to fill the agreement prefix slot.

(17)	Himu-k	si	ma	m-o-dzir-u.
	S/He-ERG	you.NOM	me.DAT	1 IO-Appl-show-3sgPAST
	'S/He showed you to me.'			

(18)	Ma	si	himu-s	g-o-dzir-i.
	I.ERG	you.NOM	her/him-DAT	2 DO-Appl-show-1sgPAST.
	'I showed you to him.'			

(19)	Ma	him	himu-s	v-o-dzir-i.
	I.ERG	her/him.NOM	her/him-DAT	1Subj-Appl-show-1sgPAST.
	'I showed him to him.'			

(20)	Himu-k	him	himu-s	Ø-o-dzir-u.
	S/he-ERG	her/him.NOM	her/him-DAT	3Subj-Appl-show-3sg PAST.
	'S/He showed him to him.'			

Sentences (13)-(20) show that verbal prefix agreement is obtained through a cyclic fashion in Pazar Laz. Within this cyclic agreement process, a hierarchy of 1/2IO > 1/2DO > 1/2Subj > 3IO/DO/Subj agreement is observed (See Öztürk & Pöchtrager [2011] for details). Nevertheless, this is not the case for all instances of verbal prefix agreement.

2.5 Non-Agentive Transitive Verbs with Dative Experiencers

The following patterns seem not to abide by the cyclic agreement observed in sentences (13) – (20).

2.5.1 Dative Experiencer-oriented Pattern

In this pattern, dative argument governs the agreement prefix (slot II). Nominative theme, acting as DO does not agree at all. Dative argument in such sentences behave as if it is an indirect object (Harris 1982) and fills the prefix position with its relevant object agreement marker (m-set) rather than the subject agreement marker (v-set). Agreement suffix is always in 3rd person, but it can reflect the plurality of the dative experiencer. This indicates the absence of a thematically marked structural subject. The default 3rd person agreement might imply the presence of a covert expletive subject, fulfilling the structural subject role. See Öztürk (2008, 2010) for a classification of different subject types in Pazar Laz.

(21) Himu-s ma g- **Ø** - o- chondr-u.
S/he-DAT I.NOM PV- **Ø**-Appl-forget-3sgPAST
'S/He forgot me.'

(22) Sk'u si/hini go-**m**-o-chondr-es
We/DAT you/they.NOM PV-**1 Obj**-Appl-forget -3sgPAST-pl
'We forgot you/them.'

(23) Si sk'u/hini go-**g**-o-chondr-u
You.DAT(sg.) we/they.NOM PV-**2 Obj**-Appl-forget-3sgPAST
'You forgot us/them.'

(24) Hini-s him/ma/hini g- **Ø** -o-chondr-es
They-DAT she/I/they.NOM PV- **Ø**-Appl-forget-3plPAST
'They forgot us/you/them.'

2.5.2 Nominative Theme-Oriented Pattern[4]

This is the pattern where nominative theme is focused. When the dative argument is deficient (3rd person) and the nominative theme DO is 1st or 2nd person then it is possible for the DO to fill Slot II. In this case dative argument does not agree at all. The DO behaves as if it is the subject and fills the prefix position with its relevant subject agreement marker (v-set) rather than the object marker (m-set). Agreement suffix can then also reflect the features of the DO.

(25)	Himu-s	ma	go-v-o-chondr-i
	S/he-DAT	I.NOM	PV- 1Subj-appl-forget-1sgPAST
	She forgot me. (me is focused)		

3 Cyclic Agreement

Anderson (1992) argues that Georgian, a language of the same family as Pazar Laz, displays a cyclic agreement on verbs. Agreement happens in a cyclic way through a list of arguments whose Morphosyntactic Representation is:

(26) $[T/A, F_{SBJ} [F_{IO} [F_{DO}]]]$

In this model, the agreement cycle starts from the innermost layer and goes towards the outer layers until the slot is filled with an agreement morpheme. In cases when an argument is deficient in terms of agreement, it leaves a dummy element '∅'. Then the agreement cycle moves further to find an agreeing argument that will fill the relevant slot.

Based on figure (26) T/A which is a verb screeve, carrying basic verb and tense meaning, starts the agreement from the DO. When DO is deficient, '∅' is inserted and then the chance to fill the relevant slot moves to IO. In cases when IO is deficient too, finally, subject agrees with the verb. In this model, DO agreement is compulsory whereas, the rest is optional. Anderson's obligatory DO rule is:

> Copy features and referential index from a direct object NP to the verb if present; if there is no Direct Object, add '∅'
> (Anderson 1992)

The most relevant point of this model to the current paper is the insertion of

[4] This pattern is beyond the scope of the current paper since it has a focused reading and one needs to determine where the arguments are on the syntactic tree when focused. See Öztürk (2011) for a syntactic account of these constructions, where it is argued that DO actually acts as the structural subject of the sentence.

dummy placeholders in cases when an argument is deficient. This suggests that one slot can be filled by placeholders and morphemes at the same time.

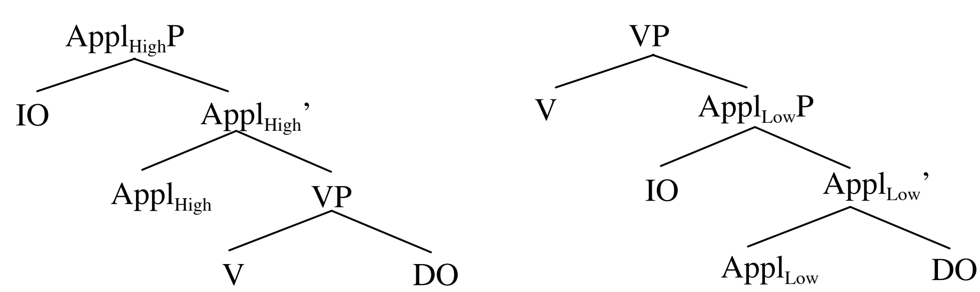
In section 1 we argued that Pazar Laz has a templatic morphology based on Inkelas (1993). The main rationale for such a claim is the complementariness of some morphemes despite semantic compatibility. Now, bringing Pazar Laz facts and Anderson's (1992) claims on cyclic agreement together, we observe a problem. Anderson's claim of inserting dummy placeholders does not agree with the fact that once a slot is filled with a morpheme it cannot host another one. If a dummy placeholder is inserted in cases of deficient arguments, then it should block other arguments from kicking in, since there is room for only one agreement marker. As a consequence we, like Anderson (1992), argue that agreement happens through a list of arguments, but unlike him, put forth that agreement slots remain empty until they are filled with the agreement marker of a non-deficient argument.

Once we pose such a claim, we come across with some problem sentences like (21), where 1st person theme is not deficient but cannot fill slot II with its relevant agreement marker. In order to suggest a solution to this problem we refer to Bejar & Rezac's (2009) cyclic agreement approach to Georgian.

Bejar and Rezac (2009) suggest a cyclic agreement model for Georgian based on Distributed Morphology. In this model, Agreement head searches the local domain first and then expands its domain towards the Spec of higher phrase in the subsequent cycles. This approach, together with Pylkännen's (2000) applicative argument, explains the cyclic agreement in Pazar Laz verbal agreement prefixes.

Pylkännen (2000) argues for the existence of two types of applicative constructions, which are High applicatives and Low applicatives. High applicatives denote a relation between an event and an individual, selecting a VP as their complement. Low applicatives, on the other hand, denote a relation between two individuals, taking DPs as their complements.

(27) a) **High Applicative** b) **Low Applicative**



(McGinnis 2001)

Öztürk (2011) and Demirok (2011), both independently, argue that Pazar Laz possesses both types of applicative constructions. Ditransitives qualify as low applicatives, whereas sentences with dative experiencers as in (21) are high applicatives (Öztürk 2011, Demirok 2011).

In the light of Bejar & Rezac (2009)'s cyclic agreement approach and Öztürk (2011) and Demirok (2011)'s applicative analyses, we propose the following agreement rule and discuss the relevant data based on this rule.

Rule Agree: In cases when there is an applicative construction, applicative head takes care of the agreement. Applicative head probes for an agreeing argument within its local domain first, then extends its domain to higher phrases in cases of deficiencies.

3.1 Ditransitives

Sentences (28)-(31) and figure (32) illustrate and support the rule proposed above.

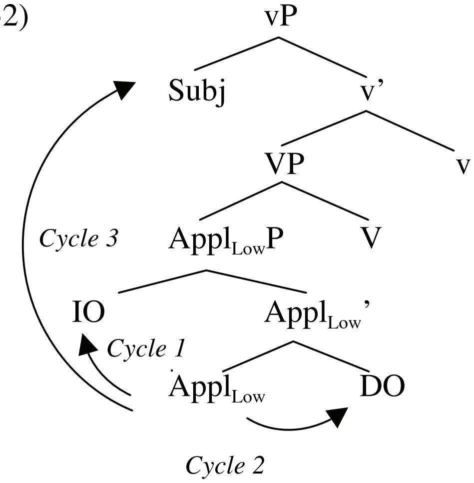
(28) **m-o-**dzir-u.
1 IO-Appl-show -3sgPAST
'He showed you to me.'

(29) **g-o**-dzir-i.
2 DO-Appl-show-1sgPAST .
'I showed you to him.'

(30) **v-o**-dzir-i.
1Subj-Appl-show-1sgPAST .
'I showed him to him.'

(31) **Ø –o**-dzir-u.
Ø -**Appl**-show-3sgPAST .
'S/He showed him to him.'

(32)



In ditransitive constructions, we assume that Appl_{Low} head takes care of the agreement in Slot II. In the first cycle, Appl_{Low} head looks at it specifier and agrees with the argument on this node, which is an IO in Pazar Laz. Sentence (28) is an example of Cycle 1 agreement. When the IO is deficient, then Cycle 2 happens and Appl_{Low} head agrees with DO as in sentence (29). When both IO and DO are deficient, Cycle 3 expands the agreement domain towards the specifier of higher phrase which is the subject in this case (Sentence (30)). Sentence (31) illustrates the cases when all the arguments are deficient.

Öztürk (2011) claims that applied arguments in Spec, ApplPs in Pazar Laz check inherent dative case. Therefore, the reason why the Appl_{Low} head first searches for the argument in its Spec, is to ensure a case and agreement match. That is, Appl head agrees with the argument whose case it checks. However, when the argument in its Spec is not compatible for agreement then it looks down for another argument within its domain, which then leads to a case and agreement mismatch in the same lines as Bhatt (2005).

3.2 Dative Experiencer – Oriented Pattern

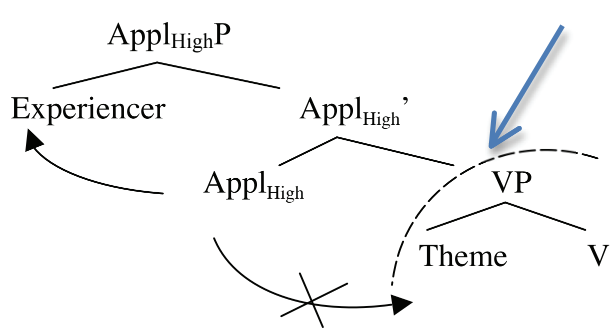
Dative experiencer-oriented pattern seem a bit problematic at first sight, since slot II remains empty even though theme is not deficient. In such sentences, only dative experiencer can fill slot II.

(33)	Sk'u	si/hini	go-**m**-***o***-chondr-es
	We.DAT	you/they.NOM	PV-**1 Obj - Appl**-forget -3plPAST
	'We forgot you/them.'		

(34)	Hini-s	him/ma/hini	g- **Ø** -***o***-chondr-es
	They-DAT	she/I/they.NOM	PV- **Ø-Appl**-forget-3plPAST
	'They forgot us/you/them.'		

Sentences with dative experiencer-oriented pattern are high applicative constructions and have the following structure:

(35)



We assume that Appl$_{High}$ head is responsible for agreement in high applicative constructions. It probes for an agreeing argument in its specifier. When the argument in its specifier is deficient, it probes down for another potential argument, which is the theme in this case. Nevertheless, theme is not in the local domain of Appl$_{High}$ head. VP functions as a blocking node. Therefore, we argue that such sentences do not pose any threat to the validity of the agreement rule we propose.

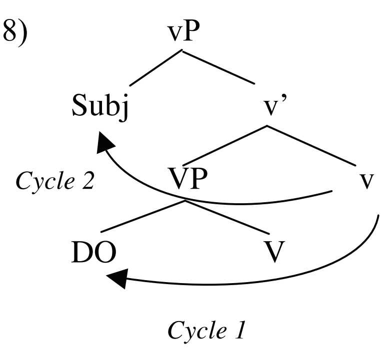
3.3 Agentive Transitives

Cyclic agreement in agentive transitives happens in a similar way with some differences. The first difference is that there is no applicative phrase in agentive transitive sentences. Therefore we assume that *v* head is responsible for the agreement. The second difference is related to the order of the cycles. In applicative constructions, first cycle would start with checking the specifier; whereas, in agentive transitive constructions this priority belongs to the DO which sits on a complement position as opposed to the subject which occupies the specifier of vP. It should also be noted that VP does not block the agreement between v head and theme, while it does so in Appl-theme agreement.

(36)	Ma	si	ce-**k**-ç-i.
	I.ERG	you.NOM	PV-**2Obj**-beat -1sgPAST
	'I beat you.'		

(37)	Ma	him	ce-**p**-ç-i.
	I.ERG	him/her.NOM	PV-**1Subj**-beat -1sgPAST
	'I beat him.'		

(38)



We propose that the reason for such an ostensible inconsistency is due to case relations between these heads and arguments. The case checking functional head has the priority in terms of agreement. Applicatives check the inherent case of the argument in their Spec position, therefore in their first cycle they first try to agree with the NP in their Spec and move on to other cycles if that NP is deficient. Similarly, v head checks the case of the theme and tries to agree with it in the first cycle. Case relation between v and theme also resolves the blocking problem caused by VP. This implies that case and agreement can be dissociated from one another under certain conditions in Laz in the same line as Bhatt (2005). See Öztürk (2011) for details of case checking in Pazar Laz.

3.4 Nominative Theme-Oriented Pattern

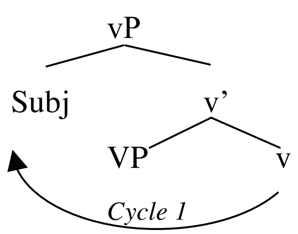
Our proposal seems to hold based on Öztürk (2011), where it is claimed that, in nominative theme-oriented pattern, direct objects act as structural subjects when focused. Focused theme raises to a node where it is checked case by T.[5] As a result of this raising, it escapes the VP blockage and enters into agree relation with the applicative head which is not possible in dative experiencer-oriented pattern. Since it case checks with T, the agreement slot II is filled by a v-set marker.

3.5 Intransitives

In intransitives, v head agrees with the sole argument using v-set agreement markers.[6] Since there is only one argument there is only one cycle for agreement.

(39)	Ma	**v**-inçir-i
	I.ERG	**1Subj**-swim -1sgPAST
	'I swam.'	

(40)



5 This node could be Spec,v or one of the multiple specifiers of the applicative head. Nevertheless, as stated earlier, this is beyond the scope of the current paper and requires further research.

6 Öztürk 2011 proposes that v-set reflects presence of case-checking with T head, whereas m-set indicates case-checking with the functional heads below vP, such as with applicatives or the v head.

4 Conclusion

In this paper, we analyzed Pazar Laz verbal agreement prefixes and put forth that cyclic agreement is compatible with templatic morphology based on a distributed morphology point of view. Agreement slots remain empty until they are filled by an agreement marker of a non-deficient argument. Another important proposal of the current paper is that in the first cycle, an agreement head tries to agree with the argument it checks case with. Pazar Laz agreement system also shows that morphology is shaped depending on syntax (argument structure, deficiencies in arguments, etc.) and therefore we claim that morphology is post-syntactic as assumed within the Distributed Morphology framework (Halle and Marantz 1993, 1994 and the following work).

References

Anderson, S. R. 1992. *A-Morphous Morphology*. Cambridge [England] ; New York: Cambridge University Press.

Bejar, S. & Rezac, M. 2009. Cyclic Agree. *Linguistic Inquiry,* Vol 40 (pp. 35-73). MIT Press.

Bhatt, R. 2005. Long Distance Agreement in Hindi-Urdu. *Natural Language and Linguistic Theory*, 23.4:757-807.

Demirok, Ö. F. 2011. A Minimalist Analysis of Laz-Georgian Applicatives. Ms. Boğaziçi University.

Emgin, B. 2009. Complementation patterns in Pazar Laz. MA thesis. Boğaziçi University.

Halle, M. and A. Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In K. Hale and S. Jay Keyser (eds.), *The View from Building 20; Essays in Honor of Sylvain Bromberger*, 111-176. MIT Press.

Halle, M. and A. Marantz. 1994. Some Key Features of Distributed Morphology. In A. Carnie, et. Al (eds.), *Papers in Phonology and Morphology*, 275-288. MITWPL 21.

Harris, A. C. 1981. *Georgian Syntax: A study in relational grammar*. Cambridge University Press.

Harris, A. C. 1982. Georgian and the Unaccusative Hypothesis, *Language* 58: 290-306.NY.

Holisky, D. A. 1991. Laz. In A. C. Harris (Eds.), *The Indigenous Languages of the Caucasus. Vol. 1. The Kartvelian Languages*. (pp. 395-472). New York: Caravan Books.

Inkelas, S. 1993. Nimboran Position Class Morphology. *Natural Language and Linguistic Theory*, 559-624.

Kojima, G., Bucaklişi, İ.A., and Uzunhasanoğlu, H. 2003. *Lazuri Grameri*. (Explanations in English G. Kojima, Eylem Bostancı). İstanbul: Chiviyazıları

McGinnis, M. 2001. Phases and the syntax of applicatives. NELS 31.

Öztürk, B. 2008. The Loss of Case System in Ardesheni Laz and its morphosyntactic consequences. Talk given at Morphological Variation and Change in the Languages of the Caucasus. 13th International Morphology Meeting. University of Vienna. Austria.

Öztürk, B. 2010 Subjects in Pazar and Ardesheni Laz. *Dilbilim Araştırmaları*. Boğaziçi University Press.

Öztürk, B. 2011 Low, High and Higher Applicatives: Evidence from Pazar Laz. Talk given at the 21st Colloquium on Generative Grammar, University of Seville, Spain.

Öztürk, B. and M. Pöchtrager (eds.). 2011. *Pazar Laz*. LINCOM Europa.

Pylkkänen, L. 2002. Introducing Arguments. Ph.D. Thesis. MIT. Cambridge, MA.

Spencer, A. 1991. *Morphological Theory: an introduction to word structure in generative grammar* . Oxford, UK ; Cambridge, Mass., USA: Basil Blackwell.

Stump, G. 2006. Template Morphology. In K. Brown (Eds)., *Encyclopedia of Language and Linguistics,* Vol 12 (pp. 559-563). Oxford: Elsevier.

Ümit Atlamaz
18 Seminary Place
New Brunswick, NJ 08901-1184

umit.atlamaz@rutgers.edu

Alignment Typology, Reflexives, and Reciprocals in Tsezic Languages

BERNARD COMRIE[1,2], DIANA FORKER[3], and ZAIRA KHALILOVA[1]
Max Planck Institute for Evolutionary Anthropology[1]; University of California Santa Barbara[2]; University of Bamberg[3]

1 The Tsezic Languages

The Tsezic (Didoic) languages form a well-defined sub-group within the Nakh-Daghestanian (East Caucasian, Northeast Caucasian) language family. They are spoken primarily in the west of the Republic of Daghestan in the Russian Federation, close to the border with Georgia, although there are also some recent settlements in lowland Daghestan and across the border in Georgia. Five individual Tsezic languages are usually recognized, although the differences within Khwarshi between Khwarshi Proper and Inkhoqwari are perhaps sufficient to consider these two distinct languages; our own Khwarshi data in this article are from the Kwantlada subdialect of Inkhoqwari. The languages, listed in what follows from north to south and then from west to east, are divided into West Tsezic – Khwarshi [khv], Tsez (Dido) [ddo], and Hinuq [gin] – and East Tsezic – Bezhta (Kapuchi) [kap] and Hunzib [huz]. Hinuq, sandwiched between Tsez and Bezhta, sometimes patterns with East rather than West Tsezic.

The Tsezic languages are all predominantly, though not rigidly, verb-final at the clause level, and more generally head-final at the phrasal level. They have a gender (noun class) system, with four or five genders depending on language and dialect. The genders are identified in examples by means of roman numerals; in all languages, gender I comprises all and only nouns with male human denotation, while gender II includes (and in some languages is limited to) all nouns denoting female humans. In the plural, only a two-way distinction is made, either human versus non-human or virile (male human) versus non-virile.

Except where more specific sources are given in the text, our data on Khwarshi are taken from Khalilova (2009), on Tsez from fieldwork by Bernard Comrie and Maria Polinsky, on Hinuq from Forker (2011), on Bezhta from fieldwork by Bernard Comrie, Madzhid Khalilov, and Zaira Khalilova (the last two also native speakers of Bezhta), and on Hunzib from van den Berg (1995).

2 Clause Structure Types

In discussing the clause structure of Tsezic languages, it is useful to identify a number of clause structure (valency) types, the most important for our present purposes being intransitive, transitive, and affective, all three of which are distinguished in parallel fashion in all Tsezic languages.

Intransitive clauses have a single core argument (where necessary abbreviated S) in the Absolutive case. If the verb can show gender–number agreement (see section 3.2 below), then it will agree with this single core argument.

Khwarshi

(1)	hadam	b-odo-še	b-eč-un.
	people(HPL.ABS)	HPL-work-IPFVCVB	HPL-be-PSTUNW
	‘The people have been working.’		

Bezhta

(2)	kid	y-e^{n}ƛ’e-yo.
	girl(II.ABS)	II-go-PST
	‘The girl went.’	

Transitive clauses have two core arguments, one typically more agent-like (and abbreviated A) in the Ergative case, the other typically more patient-like (abbreviated P) in the Absolutive case. If the verb can show gender–number agreement, then it agrees with the P argument.

Khwarshi

(3)	heč’č’e	atɣul	madinat-i	ō^{n}ču	b-ez-un.
	most	in.front	Madinat(II)-ERG	hen (III.ABS)	III-buy-PSTUNW
	‘First Madinat bought the hen.’				

Bezhta

(4)	mexanik-li	radio	y-it’il-lo.
	mechanic(I)-ERG	radio(IV.ABS)	IV-repair-PST
	‘The mechanic repaired the radio.’		

Affective clauses contain verbs expressing perceptions, emotions, etc. In the affective clause there are again two core arguments, an experiencer-like argument (abbreviated Exp) in the Lative case and a stimulus-like argument (abbreviated Stim) in the Absolutive case. If the verb can show gender–number agreement, then it agrees with the Stim argument. It will be noted that the general rule for verb indexing in the Tsezic languages is that verbs agree only with their core argument in the Absolutive case.

Khwarshi

(5)	bet'erhan-ɨl	b-ak-un	boc'o.
	owner(I)-LAT	III-see-PSTUNW	wolf(III.ABS)
	'The owner saw the wolf.'		

Bezhta

(6)	di-l	kid	y-ac-ca.
	me(I)-LAT	girl(II.ABS)	II-love-PRS
	'I (male speaker) love the girl.'		

There are also other, sometimes language-specific clause types that will not play any major role in what follows. For instance, example (7) illustrates the potential construction in Hinuq, in which the most agent-like argument stands in the At-essive case (literally expressing location at), the typically patient-like argument in the Absolutive.

Hinuq

(7)	łac-qo	ac	y-aɣi-ł-o	gom.
	wind-AT.ESS	door(IV.ABS)	IV-open-POT-IPFVCVB	be.NEG
	'The wind can't open the door.'			

3 Morphological Alignment

The main morphological phenomena relevant to alignment typology in Tsezic languages are case marking, discussed in section 3.1, and verb indexing, discussed in section 3.2.

3.1 Case Marking

The Tsezic languages, like most other Nakh-Daghestanian languages, have rich case inventories, composed primarily of spatial cases, but only a small number of cases are relevant for present purposes. All of intransitive S, transitive P, and affective Stim stand in the Absolutive case. Transitive A stands in the Ergative case, while affective Exp stands in the Lative case. Alignment thus groups together S, P, and Stim as opposed to A or Exp. In comparing intransitive and transitive clauses, this is ergative-absolutive alignment of case marking.

The Absolutive case is always identical to the citation form of the noun phrase in question. The Lative case is always distinct from the Absolutive case (and, indeed, all other cases) through the presence of the Lative suffix. The morphology of the Ergative is more complex. For most noun phrases in most Tsezic languages, the Ergative is distinct from the Absolutive (and all other cases), sometimes through the presence of a distinct Ergative suffix, sometimes through the use of an Oblique stem distinct from the Absolutive, with the Ergative having no additional suffix, all other oblique cases using the Oblique stem plus a case suffix. The distribution of these two (and occasional other) types varies from language to language, and even from noun phrase to noun phrase within a language. To this general pattern of an Ergative case distinct from the Absolutive there are two exceptions.

The first concerns first- and second-person pronouns. Only in Khwarshi do all these personal pronouns have distinct Ergative and Absolutive cases. In Tsez, this case distinction is made in the plural, but not in the singular. In Hinuq and the East Tsezic languages, no first- and second-person pronouns make this case distinction. The relevant forms are set out in (8).

(8) First- and second-person pronouns

	1SG		2SG		1PL		2PL	
	S/P	A	S/P	A	S/P	A	S/P	A
Khwarshi	do	de	mo	me	íl^{y}o	il^{y}é	mížo	mižé
Tsez	di	di	mi	mi	eli	elā	meži	mežā
Hinuq	de	de	me	me	eli	eli	meži	meži
Bezhta	do	do	mi	mi	ile	ile	miže	miže
Hunzib	də	də	mə	mə	ile	ile	miže	miže

The second concerns only Bezhta. In this language, the Ergative is identical to the Oblique stem, so for nouns with an Oblique stem distinct from the Absolutive, the Ergative is distinct from the Absolutive. However, a good number of nouns have identity of Absolutive and Oblique stems, and thus of Absolutive and Ergative cases. Some examples are given in (9). It should be noted that whether a noun has or lacks an Ergative/Absolutive distinction does not correlate with such features as animacy, which often controls similar patterns in other languages; see further Comrie (2001a).

(9) Selected case forms in Bezhta

	ABS SG	OBL stem	ERG SG	LAT SG
'brother'	is	ist'i-	ist'i	ist'i-l
'sister'	isi	isi-	isi	isi-l
'fox'	sora	sorali-	sorali	sorali-l
'horse'	soyya	soyya-	soyya	soyya-l
		soyyali-	soyyali	soyyali-l
'forest'	wan	wana-	wana	wana-l
'river'	e^{n}xe	e^{n}xe-	e^{n}xe	e^{n}xe-l

3.2 Verb Indexing

In the Tsezic languages, nearly all vowel-initial verbs index (agree with) one of their arguments in gender–number by means of a single-consonant prefix. A few vowel-initial verbs do not show agreement, perhaps reflecting an earlier stage where there was an initial consonant since lost. Verbs beginning with ʕV sometimes also show agreement in the same way, the sequence perhaps to be analyzed phonologically as a pharyngealized vowel. In some Tsezic languages, a few verbs also show agreement by means of internal vowel change.

By way of illustration, the agreement prefixes of Tsez are shown in (10). Note that gender I has a zero prefix – this is constant across the Tsezic languages – but the absence of an overt prefix on a verb that takes agreement is always interpreted as indexing a gender I argument; it cannot be interpreted as any kind of missing or default agreement.

(10) Tsez agreement prefixes on verbs

SG				PL	
I	II	III	IV	I	non-I
Ø-	y-	b-	r-	b-	r-

As noted in section 2, where a verb can index an argument, this is always the S, P, or Stim, never the A or Exp. In comparing intransitive and transitive clauses, again we have an instance of ergative-absolutive alignment, agreement being only with the absolutive argument – note that this applies even in those instances where the noun phrase itself does not make a distinction between Absolutive and Ergative cases.

Given that in general only vowel-initial verbs show indexing, one might wonder whether indexing plays any significant role in practice in marking the grammatical relations of noun phrases, especially since most noun phrases distinguish Ergative and Absolutive, and all have a distinct Lative. And indeed, if one counts verbs in the lexicon, only a minority allow indexing. For Khwarshi, Khalilova (2009:181) finds that about 70% of verbs are consonant-initial and therefore cannot show indexing, a further 7% are vowel-initial but do not show indexing, while only 23% are vowel-initial and show indexing. However, it turns out that the vowel-initial verbs that show indexing include some of the most frequent verbs in the language, including some that are frequently used as auxiliaries in periphrastic constructions. Thus, the first tale in Abdulaev and Abdullaev (2010), *The Rainbow*, comprises a total of 281 words, of which 53 are verb forms showing agreement, 35 (including 8 vowel-initial) verb forms not showing agreement, i.e. forms showing agreement outnumber those not showing agreement in text by a ratio of about 3:2. Since the Tsezic languages have a tendency to omit noun phrases that are retrievable from context, rather than using pronouns, the importance of verb indexing in reference tracking is greater than might seem from a purely grammatical description of the phenomenon.

4 Syntactic Alignment

Many phenomena that provide good tests for alignment differences in a number of languages across the world do not do so in Tsezic languages. First, many syntactic phenomena in Tsezic languages are neutral with respect to alignment, for instance all major constituents of the clause are accessible to such constructions as relative clause formation and content question formation; for relative clauses in Tsez, see Comrie and Polinsky (1999). Second, there is evidence that such phenomena as pronominalization (whether by zero anaphora or overt pronouns) in Tsezic languages are governed to at least a large extent by pragmatic factors; whether and, if so, to what extent syntactic constraints are involved is a task for future research. Nonetheless, there are some phenomena that are sensitive to differences in grammatical relations and do therefore provide evidence in favor of syntactic alignment in Tsezic languages.

4.1 Control

We use "control" here in at least one of the senses current in formal grammar, namely to indicate the obligatory coreference of a missing argument of a dependent clause with an overt (or understood) argument of a matrix clause. We are concerned with the identification of the missing argument in the dependent clause. Consider Tsez examples (11)–(13) (discussed further in Comrie 2000, 2004).

Tsez

(11)	dā̄-r	new-ā-ɣor	Ø-ik'-a	r-eti-x.
	me-LAT	Mokok-IN-DIR	I-go-INF	IV-want-PRS
	'I want to go to Mokok.'			
(12)	dā̄-r	kaɣat	cax-a	y-eti-x.
	me-LAT	letter(II.ABS)	write-INF	II-want-PRS
	'I want to write a letter.'			
(13)	uži-r	kid	y-ukad-a	y-eti-s.
	boy-LAT	girl(II.ABS)	II-see-INF	II-want-PSTWIT
	'The boy wanted to see the girl.'			

In each example, the infinitive in the dependent clause lacks an argument that must be interpreted as coreferential with the Lative (experiencer) argument of the matrix verb 'to want'. In (11), the missing argument is the S of the intransitive verb 'to go'. In (12), it is the A of the transitive verb 'to write'. In (13), it is the Exp of the affective verb 'to see'. In each instance, this is the only possibility, in particular it is not possible to omit the P of a transitive verb or the Stim of an affective verb under coreference with the appropriate argument in the matrix clause. In comparing transitive and intransitive clauses, alignment is thus nominative-accusative (A treated like S, and unlike P), while in comparing affective and intransitive clauses we have likewise parallel treatment of Exp and S, different treatment of Stim, in each instance going against the morphological alignment in terms of case marking and verb indexing.

4.2 Imperative

The situation with imperatives is a little more complex. First, if we compare transitive and intransitive imperative sentence, all Tsezic languages behave like Khwarshi in examples (14)–(15), i.e. the addressee of an imperative sentence may be either the S of an intransitive or the A of a transitive, but not the P of a transitive clause. In other words, here we have nominative-accusative alignment.

Khwarshi

(14) Ø-o^{n}k'-o, obu.
I-go-IMP father(I)
'Go, father!'

(15) miže l-i-yo.
you.PL(ERG) IV-do-IMP
'You do it!'

The complications arise with affective verbs, as discussed by Comrie (2001b), though in this early article only comparing Tsez and Bezhta (which happen to represent opposite extremes), in ignorance of the more subtly differentiated picture in Khwarshi and Hinuq. As illustrated in (16), Hinuq allows an imperative to be formed where the addressee is the Exp of the verb 'to love,' i.e. parallel treatment of Exp and S, with different treatment of Stim, since it is not possible for Stim to be the addressee of an imperative sentence. Example (17) provides an alternative way of expressing essentially the same information, by causativizing the affective verb, which produces a transitive imperative sentence in which the A can, as usual, be the addressee.

Hinuq

(16) debe-z hado uži Ø-eti.
you-LAT this boy(I.ABS) I-love(IMP)

(17) me hado uži Ø-eti-r-o.
you((ERG)) this boy(I.ABS) I-love-CAUS-IMP
'You love this boy!'

By contrast, Tsez simply disallows the formation of imperative sentences from affective clauses, i.e. (18) is impossible, where the addressee is the Exp, as equally would be an example where the addressee is the Stim. Causativizing the affective verb to give a transitive verb leads to the grammatical sentence (19), where the addressee is A of the imperative sentence.

Tsez

(18)	*mežu-l	mežu-s	tušman-bi	b-eti.
	you.PL-LAT	you.PL-GEN	enemy-PL(ABS)	IPL-love(IMP)
(19)	mež-ā	mežu-s	tušman-bi	b-eti-r.
	you.PL-ERG	you.PL-GEN	enemy-PL(ABS)	IPL-love-CAUS(IMP)
	'Love your enemies!'			

In other words, at least under some circumstances and in some Tsezic languages, affective clauses simply disallow formation of an imperative sentence. More specifically: The formation of imperative sentences from affective clauses seems to be possible quite generally in Bezhta. It seems to be absolutely excluded in Tsez and Hunzib, with only the alternative causative, and therefore transitive, construction being allowed. (For the Hunzib data, see van den Berg (1995:88); note that van den Berg uses "inversive" for our "affective.") In Khwarshi, imperatives from affective verbs seem generally possible, but are excluded with 'to find' and 'to see'. In Hinuq, imperatives from affective verbs are generally not possible, only the verb 'to love', as in (16), allows this possibility. It will be noted that the extent of the possibility of imperative sentences formed from affective clauses bears no close correlation either to the genealogical division of Tsezic languages into West and East or to geographical adjacency.

4.3 Reflexives and Reciprocals

The phenomena described so far, including those relating to syntactic alignment in sections 4.1 and 4.2, are perhaps not too surprising from a cross-linguistic perspective, even if the details of the Tsezic languages at times provide interesting language-specific variations on a universal theme. With respect to control and imperatives, for instance, Dixon (1994:131–137) argues that nominative-accusative alignment is effectively guaranteed on a semantic basis, and that languages would not be expected to differ in this respect, irrespective of their morphological alignment or of their syntactic alignment in constructions where such semantic factors are not present (such as omission of coreferential noun phrases in clause coordination).

Particular interest is therefore provided by reflexive and reciprocal constructions in the Tsezic languages, which do appear, either optionally or obligatorily depending on the language and precise configuration, to violate proposed universals that would favor nominative-accusative syntax in the direction of ergative-absolutive syntax, with corresponding patterns in affective clauses. It should be noted that we are concerned here only with reflexives and reciprocals where two coreferential arguments, one identifiable as a reflexive or reciprocal marker, are present in the morphosyntactic structure of the construction. Constructions where reflexivity or reciprocality is expressed by means of reducing the valency of the predicate are irrelevant to the issues at hand.

Dixon (1994:138–139) continues his discussion by saying, with respect to reflexives, that “in every ergative language, as in every accusative language, the ‘antecedent’, i.e. the controller of reflexivity is A.” In similar vain, Haspelmath (2007:2096) says with respect to reciprocals that “less prominent arguments cannot antecede more prominent arguments.” (From the context, it is clear that for Haspelmath A is more prominent than P.) We may refer to patterns where the more prominent argument antecedes the less prominent one as “canonical” reflexives or reciprocals.

Of course, in order to test such claims it is necessary to have a more precise notion of “prominence”, but it is clear from sources such as those cited that A will be more prominent than P, probably that Exp will be more prominent than Stim, and that in a language with a well-defined category of subject then subject will be more prominent than other grammatical relations. In languages with “promotional” voice systems, the hierarchy placing A above P or Exp above Stim will sometimes conflict with that placing subject above object, for instance in passive constructions, so one might expect to find some cross-linguistic variation here, although where A/Exp is subject and P/Stim is non-subject, then the prominence relation is clear. In this respect, it is useful to compare English and Tagalog. (The English data below include both reflexives and reciprocals; the Tagalog data include only reflexives, since in Tagalog reciprocals involve detransitivization, i.e. they do not retain A and P as distinct arguments.)

In English example (20), both the A > P and the subject > non-subject hierarchies are maintained, with the antecedent *John* (A and subject) and the anaphor *himself* (P and object), and this is indeed the only fully acceptable example from the quadruple. Example (21) violates both hierarchies, and is completely unacceptable. Version (22) violates the A > P hierarchy (in terms of the lexical arguments of the verb ‘to hit’), but not the subject > non-subject one, and is marginal. Version (23) violates the subject > non-subject hierarchy, but not the A > P hierarchy, and is completely unacceptable. It seems that in English the hierarchy subject > non-subject is absolute with regard to reflexivization, with the A > P hierarchy playing a less significant role.

(20) John hit himself.
(21) *Himself hit John.
(22) ?John was hit by himself.

(23) *Himself was hit by John.

Examples (24)–(27) show that the same holds for reciprocals in English.

(24) John and Mary hit each other.
(25) *Each other hit John and Mary.
(26) ?John and Mary were hit by each other.
(27) *Each other were hit by John and Mary.

In Tagalog (here following essentially Schachter 1977:292–293) again, two levels of representation can be recognized, one corresponding to the lexical argument structure of the predicate, in which A > D (or more generally: A > non-A), the other corresponding to a voice-like distinction – “focus”, in traditional Philippinist terminology – where F > non-F. (The noun phrase selected as F is marked by the preposed particle *ang*, and the verb form encodes whether A or D has been selected as F. The English translations are necessarily approximate and sometimes marginal or even unacceptable.) In examples (28) and (29), the A > D hierarchy is maintained, while in (30) and (31) it is violated. In (28) and (31), the F > non-F hierarchy is maintained, while in (29)–(30) it is violated. Clearly, in Tagalog only the A > D hierarchy is relevant to reflexivization, which noun phrase is selected as F is irrelevant.

Tagalog

(28)	Nag-aalala	ang	lolo	sa	kaniyang_sarili.
	AF-worries	F	grandfather	D	REFL
	‘Grandfather worries about himself.’				
(29)	In-aalala	ng	lolo	ang	kaniyang_sarili.
	DF-worries	A	grandfather	F	REFL
	‘Himself is worried about by Grandfather.’				
(30)	*Nag-aalala	sa	lolo	ang	kaniyang_sarili.
	AF-worries	D	grandfather	F	REFL
	‘Himself worries about Grandfather.’				
(31)	*In-aalala	ang	lolo	ng	kaniyang_sarili.
	DF-worries	F	grandfather	A	REFL
	‘Grandfather is worried about by himself.’				

Before turning to Tsezic data, it is worth noting that potential exceptions to generalizations like those proposed by Dixon and Haspelmath have been noted in the earlier literature, more specifically for West Caucasian (Northwest Caucasian) languages. Thus, Smeets (1984:268) analyzes the Adyghe (West Circassian) [ady] reciprocal construction as in (32) as having the reciprocal prefix in the A slot and the first person plural prefix in the P slot, which would mean that the antecedent is P and the reciprocal A.

Adyghe
(32) tə-zere-ɬeɣʷə-ɣ.
1PL-RECIP-see-PST
'We saw each other.'

This contrasts, incidentally, directly with the corresponding reflexive, which would have the canonical pattern, as in (33), where the reflexive prefix is in the P slot, the antecedent in the A slot.

Adyghe
(33) zə-t-ɬeɣʷə-ɣ.
REFL-1PL-see-PST
'We saw ourselves.'

However, Kazenin (2007:751), writing on the closely related Kabardian (East Circassian) [kbd], while placing the reciprocal prefix in the A slot, nonetheless considers that it detransitivizes the verb, i.e. (32) would be an intransitive reciprocal construction and the question of a prominence relation between two arguments would not arise. Letuchiy (2007:809), again dealing with Adyghe, cites examples where both arguments are independent words, the antecedent being a noun phrase preferably in the Oblique case (which subsumes A), i.e. a canonical reciprocal, but with a less preferred, questionable alternative where it stands in the Absolutive (subsuming P).

Clearly, the data on West Caucasian reciprocals are complex and have been subject to different analyses, some but not all of which would make them non-canonical. Moreover, since the main construction seems to involve prefixes within the verb morphology rather than independent noun phrases, apparent violations of canonicity might be put down to the vagaries of morphology. In any event, more work is needed on reciprocals in these languages.

Fortunately, the data from the Tsezic languages are clear, both for reflexives and for reciprocals, and provide clear evidence of violations of canonicity. Nonetheless, we need to make some caveats with respect to the following data. Reflexive and reciprocal constructions are extremely rare in the texts to which we have had access, and most of the following data are therefore elicited. Where possible, and this applies especially to the Khwarshi, Hinuq, and Bezhta data, we have made the usual attempts to ensure that our data are as reliable as possible, including checking with multiple speakers. In some instances, we also tried to elicit alternative constructions to the version initially offered, in particular to test variations in word order. We consider these data, especially on alternative word orders, less reliable, but have nonetheless given the judgments of our consultants where we have them. The Tsez data are more restricted, in that they relate to reflexives but not (with one exception) to reciprocals, and are taken from Polinsky and Comrie (2003). We have so far no comparable data for Hunzib.

4.3.1 Reflexives in Tsezic

In Tsez transitive constructions, the relation between antecedent and anaphor is canonical, with the A as antecedent and the P as anaphor, as in (34). The equivalent non-canonical relation of antecedent and anaphor is judged unacceptable. However, inverting the word order in (34) to place the anaphor before the antecedent is judged acceptable. In the affective construction, only the non-canonical construction, as in (35), where the antecedent is the Stim (in the Absolutive) and the anaphor the Exp (in the Lative) is accepted; changing the order of the two noun phrases is very questionable.

Tsez

(34)	ʕal-ǟ	nesǟ_že	žek’-si.
	Ali(I)-ERG	REFL(I.ABS)	hit-PSTWIT
	‘Ali hit himself.’		
(35)	pat’i	neło<r>_že	y-eti-x.
	Pati(II.ABS)	REFL(II)<LAT>	II-love-PRS
	‘Pati loves herself.’		

Bezhta has two different reflexive formations available clause-internally, which have different properties. The first is the simple reflexive, such as Absolutive *žu* in (36) and (38). With this choice of reflexive, the construction is canonical, with the A or Exp as antecedent, the P or Stim as anaphor. Changing the order of the two arguments is not permitted. (An instance of *žu* preceding its intended anaphor is interpreted as coreferential with an antecedent in a previous sentence, indicating topic continuity.) The compound reflexive, as in (37)–(39), reverses the antecedent-anaphor relation, since now the P or Stim is antecedent, the A or Exp anaphor. If the compound reflexive is used, then the linear order of antecedent and anaphor may be inverted.

Bezhta

(36)	murad-i	žu	Ø-uɣo-l-lo.
	Murad(I)-ERG	REFL(ABS)	I-die-CAUS-PST
(37)	murad	hinis_hin-i	Ø-uɣo-l-lo.
	Murad(I.ABS)	REFL-ERG	I-die-CAUS-PST
	‘Murad killed himself.’		
(38)	ist’i-l	žu	Ø-ac-ca.
	brother(I)-LAT	REFL(ABS)	I-like-PRS

(39)	is	hinis_hini-l	Ø-ac-ca.
	brother(I.ABS)	REFL-LAT	I-like-PRS
	'Brother likes himself.'		

Khwarshi, like Tzez, has only a compound reflexive. In both transitive and affective constructions, it allows both canonical (as in (40), (42)) and non-canonical (as in (41), (43)) relations between antecedent and anaphor. In all four examples, inverting the linear order of antecedent and anaphor is judged acceptable.

Khwarshi

(40)	ražab-i	žu_žuč	Ø-uwox-i.
	Rajab(I)-ERG	REFL(ABS)	I-kill-PSTWIT
(41)	ražab	ise_iseč	Ø-uwox-i.
	Rajab(I.ABS)	REFL.ERG	I-kill-PSTWIT
	'Rajab killed himself.'		
(42)	musa-l	žu_žuč	Ø-iyōq'.
	Musa(I)-LAT	REFL(ABS)	I-know.GNT
(43)	musa	ise_isu-l	Ø-iyōq'.
	Musa(I.ABS)	REFL-LAT	I-know.GNT
	'Musa knows himself.'		

For transitive constructions, Hinuq allows either a simple reflexive, as in (44), or a compound reflexive, as in (45). In both cases, the relation between antecedent and anaphor is canonical. (Inversion of the linear order of antecedent and anaphor remains to be checked for (44); for (45), it is disallowed.) For affective constructions, there are two possibilities. Sentence (46) illustrates the analytically more straightforward of these, with a compound reflexive, and with the non-canonical relation between antecedent and anaphor. Inverting the linear order of antecedent and anaphor is possible.

Hinuq

(44)	maħama-y	zo	Ø-uher-iš.	
	Mahama(I)-ERG	REFL(ABS)	I-kill-PSTWIT	
	'Mahama killed himself.'			
(45)	šayix-i	zoni_zo	zok-ko.	
	Sheikh-ERG	REFL(ABS)	beat.prs	
	'Sheikh beats himself.'			
(46)	madina	zoni_zon-ez	y-eq'i-yo.	
	Madina(II.ABS)	REFL-LAT	II-know-PRS	
(47)	madina-z	zon-ez	zo	y-eq'i-yo.
	Madina(II)-LAT	REFL-LAT	REFL(ABS)	II-know-PRS
	'Madina knows herself.'			

Example (47) is more complex. The antecedent *madinaz* is in the Lative case, the case appropriate to Exp, which suggests a canonical relation between antecedent and anaphor. The problem is the “compound reflexive’ *zonez zo*. Compound reflexives of this type, where the first component echoes the case of the antecedent, are found only in affective and a few other constructions – for instance, the potential illustrated in (7) – which makes it difficult to generalize. However, inverting the linear order of antecedent and anaphor is possible, giving *zonez zo madinaz yeq’iyo*, where the structure *zonez zo* is treated as a single unit for movement purposes. We therefore assume that *zonez zo* is some kind of Absolutive of the reflexive pronoun, and treat (47) as an instance of the canonical relation between antecedent and anaphor.

4.3.2 Reciprocals in Tsezic

In this section we start with Bezhta data, which provide the clearest counterexamples to the canonical relation between antecedent and anaphor. In both (48) and (49), illustrating transitive and affective clauses respectively, the only possible relation is where the P or Stim is the antecedent and the A or Exp the anaphor. In both examples, inverting the linear order of antecedent and anaphor is possible.

Bezhta

(48)	kid-na	öžö-nä	sid<i>_hos	b-iyaƛ’e-yo.
	girl(ABS)-and	boy(ABS)-and	RECIP<ERG>	IPL-kill.PL-PST
	‘The girl and the boy killed each other.’			
(49)	pat’imat-na	rasul-na	sidi<l>_hosso	b-āc-ca.
	Patimat-and	Rasul-and	RECIP<LAT>	IPL-like-PRS
	‘Patimat and Rasul like each other.’			

In Khwarshi, in both transitive and affective clauses, both canonical ((51) and (53)) and non-canonical ((50) and (52)) relations between antecedent and anaphor are possible, and the linear order of antecedent and anaphor can be inverted.

Khwarshi

(50)	ɣ^{ʕw}e-bo	hadiyad-za	ha^{n}ha^{n}n-i.
	dog-PL.ABS	RECIP-ERG	bite-PSTWIT
(51)	ɣ^{ʕw}e-za	hadiyad-ba	ha^{n}ha^{n}n-i.
	dog-PL.ERG	RECIP-ABS	bite-PSTWIT
	‘The dogs bit each other.’		
(52)	izzu	hadiyadi-l	goq-še.
	they(ABS)	RECIP-LAT	like-PRS
(53)	izzu-l	hadiyad-ba	goq-še.
	they-LAT	RECIP-ABS	like-PRS
	‘They like each other.’		

In Hinuq, the transitive construction allows only the canonical relation between antecedent and anaphor, as in (54), while the affective construction allows both relations, as in (55)–(56). In all three examples, the linear order of antecedent and anaphor can be inverted.

Hinuq			
(54)	haze-y	sedihes	haƛi-š.
	they-ERG	RECIP(ABS)	push-PSTWIT
	'They pushed each other.'		
(55)	hagbe	sedised-ez	b-eti-yo.
	they(ABS)	RECIP-LAT	IPL-love-PRS
(56)	hagze-z	sedihes	b-eti-yo.
	they-LAT	RECIP(ABS)	IPL-love-PRS
	'They love each other.'		

The data presented in sections 4.3.1–2 can be summarized as in tables (57)–(58). Where both canonical and non-canonical relations are possible, we have placed canonical above non-canonical in (57), as there seems to be some preference for the canonical relation where it is possible with transitive constructions. Conversely, in (58) we have placed non-canonical first, as this seems to be the preferred version where alternatives are possible with affective constructions. In both tables, non-canonical relations are boldfaced. Instances where only the non-canonical relation is possible are, of course, of particular interest.

(57) Transitive construction

	Reflexive Antecedent	Reflexive	Reciprocal Antecedent	Reciprocal
Tsez	A	P		
Bezhta	A	P		
	P	**A**	**P**	**A**
Khwarshi	A	P	A	P
	P	**A**	**P**	**A**
Hinuq	A	P	A	P

(58) Affective construction

	Reflexive Antecedent	Reflexive	Reciprocal Antecedent	Reciprocal
Tsez	**Stim**	**Exp**		
Bezhta	**Stim**	**Exp**	**Stim**	**Exp**
	Exp	Stim		
Khwarshi	**Stim**	**Exp**	**Stim**	**Exp**
	Exp	Stim	Exp	Stim
Hinuq	**Stim**	**Exp**	**Stim**	**Exp**
	Exp	Stim	Exp	Stim

5 Conclusions and Prospects

While the alignment properties of case marking, verb indexing, control, and imperatives in the Tsezic languages might be judged to provide at best minor variations on well attested themes, the alignment properties of reflexive and reciprocal constructions provide major challenges to accepted views on the canonicity of prominence hierarchies in such constructions. In particular, the Tsezic languages provide frequent instances where violations of proposed universal prominence relations between antecedent and anaphor are possible, and even some where such violations are obligatory (the reflexive affective in Tsez, both transitive and affective reciprocals in Bezhta). Clearly, a phenomenon believed to be impossible is now attested.

Now that the phenomenon is attested, future work will need to establish precisely in which languages it is encountered. It is clearly rampant in Tsezic languages. Preliminary data suggest that it may be present in some (though not all) other Nakh-Daghestanian languages, although in some cases more work needs to be done on the analysis of the construction in question (e.g. in some cases the reciprocal may be an adverb rather than a pronoun). Likewise, further analysis seems required before accepting into the fold the West Caucasian examples discussed in section 4.3. However, we are not aware of even potential candidates from other parts of the world. We seem, therefore, to have an areally highly restricted phenomenon – essentially, the North Caucasus (in part) versus the rest of the world! We know of no reason why this should be so; indeed, if non-canonical relations between antecedent and anaphor are a feature of ergative syntax, one might well have expected to find them in languages with rampant ergative syntax, rather than in languages like the Tsezic languages where there is little ergativity beyond morphology.

Of course, one must also consider the possibility that the Tsezic data should be given a different analysis. While in general we leave this as an open challenge, we will finish by suggesting one direction such a reanalysis might take. We do not think that the Tsezic non-canonical reflexive and reciprocal can be analyzed as adverbs in an intransitive construction, given that the case of the reflexive or reciprocal pronoun varies between Ergative and Lative according to what would be expected for an A or Exp in a transitive or affective clause. However, one might want to explore the possibility that non-canonical relations between antecedent and anaphor are not a syntactic phenomenon, but rather a purely morphological one, paralleling morphological ergativity in Tsezic languages as discussed in section 3. One piece of evidence in favor of this is the preference for word orders where the antecedent precedes the anaphor – even where alternatives are possible, they were normally only provided in response to an explicit question; i.e. there is still some sense in which reflexive and reciprocal constructions in Tsezic languages are canonical, namely in the linear order of antecedent before anaphor. A perhaps more striking piece of evidence comes from the interaction of reflexives and reciprocals with control phenomena as discussed in section 4.1. Take the case of Bezhta, where the non-canonical relation is obligatory in reciprocal constructions. Now imagine we want to embed ‘the girls praised each other’ under a verb of wanting to get ‘the girls want to praise each other’. In the basic structure of the dependent clause, the A will be the reciprocal in the Ergative, while the P will be the antecedent in the Absolutive. In control structures, Bezhta would normally omit the A of the dependent clause, under coreference with the overt noun phrase ‘girls’ in the matrix clause. However, what actually happens in this configuration in Bezhta is as illustrated in (59).

Bezhta

(59)	kibbā-l	y-at’-na	gey	sid<i>_hos	wecci<b>ow-al.
	girl.PL-LAT	IV-want-CVB	be.PRS	RECIP<ERG>	praise<HPL>-INF

‘The girls want to praise each other.’
[lit. ‘The girls want for each other to praise [them].’]

It is not the Ergative reciprocal pronoun that is omitted in the dependent clause, but rather its Absolutive antecedent, suggesting that the reciprocal pronoun, though Ergative, is perhaps not the A of its clause. Exploration of this and other possibilities remains a task for future research.

Transcription, Glossing, and Abbreviations

The transcription used for Tsezic languages reflects a broad transcription developed, with minor variations, by a number of scholars working on these languages and based on earlier transcriptions of Caucasian, especially Daghestanian languages. The following IPA correspondences should be noted: ä = [æ], ö = [ø], ^{n} indicates nasalization of the preceding vowel, a macron indicates a long vowel; c = [ts], č = [tʃ], š = [ʃ], ž = [ʒ], ƛ = [tɬ], y = [j], ^{y} indicates palatalization of the preceding consonant; an acute accent indicates word accent, marked only where relevant.

Glossing conventions follow the Leipzig Glossing Rules, for which see: http://www.eva.mpg.de/lingua/resources/glossing-rules.php. Abbreviations used are the following:

A	Agent-like argument	INF	Infinitive
ABS	Absolutive	IPFVCVB	Imperfective converb
AF	Actor focus	LAT	Lative (motion to)
AT	locative ‘at’	NEG	Negative
CAUS	Causative	OBL	Oblique
CVB	converb	P	Patient-like argument

D	Direction	PL	Plural
DF	Direction focus	POT	Potential
DIR	Directional	PRS	Present
DYN	Dynamic	PST	Past
ERG	Ergative	PSTUNW	Past unwitnessed
ESS	Essive (location)	PSTWIT	Past witnessed
Exp	Experiencer-like argument	RECIP	Reciprocal
F	Focus (in Philippine sense)	REFL	Reflexive
GEN	Genitive	S	Single argument of intransitive
GNT	General tense		
HPL	Human plural	SG	Singular
IMP	Imperative	Stim	Stimulus-like argument
IN	locative 'in'		

References

Abdulaev, Arsen K. and Isa K. Abdullaev. 2010. *Cezyas Folklor* [*Tsez Folklore*]. Leipzig–Makhachkala: Lotos.

Comrie, Bernard. 2000. Valency-Changing Derivations in Tsez. In R.M.W. Dixon and Alexandra Y. Aikhenvald, eds., *Changing Valency: Case Studies in Transitivity*, 360–374. Cambridge: Cambridge University Press.

Comrie, Bernard. 2001a. How Independent is Tsezic Morphology? In Mary Andronis, Christopher Ball, Heidi Elston, and Sylvain Neuvel, eds, *The Panels* (CLS 37-2), 367–383. Chicago: Chicago Linguistic Society.

Comrie, Bernard. 2001b. "Love Your Enemies": Affective Constructions in Two Daghestanian Languages. In István Kenesei and Robert M. Harnish, eds., *Perspectives on Semantics, Pragmatics, and Discourse: A Festschrift for Ferenc Kiefe*r, 59–72. Amsterdam: John Benjamins.

Comrie, Bernard. 2004. Oblique-Case Subjects in Tsez. In Peri Bhaskararao and Karumuri Venkata Subbarao, eds., *Non-Nominative Subjects*, Vol. 1, 113–127. Amsterdam: John Benjamins.

Comrie, Bernard and Maria Polinsky. 1999. Form and Function in Syntax: Relative Clauses in Tsez. In Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan, and Kathleen Wheatley, eds., *Functionalism and Formalism in Linguistics,* Vol. II: *Case Studies*, 77–92. Amsterdam: John Benjamins.

Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Forker, Diana. 2011. A Grammar of Hinuq. Ph.D. dissertation, University of Leipzig, Leipzig.

Haspelmath, Martin. 2007. Further Remarks on Reciprocal Constructions. In Vladimir P. Nedjalkov, ed., *Reciprocal Constructions*, Vol. 4, 2087–2115. Amsterdam: John Benjamins.

Kazenin, Konstantin I. 2007. Reciprocals, Comitatives, Sociatives, and Reflexives in Kabardian. In Vladimir P. Nedjalkov, ed., *Reciprocal Constructions*, Vol. 2, 739–771. Amsterdam: John Benjamins.

Khalilova, Zaira. 2009. A Grammar of Khwarshi. Ph.D. dissertation, Leiden University, Leiden. Published Utrecht: LOT, Netherlands Graduate School of Linguistics. Downloadable from http://www.lotpublications.nl/index3.html.

Letuchiy, Alexander. 2007. Reciprocals, Reflexives, Comitatives, and Sociatives in Adyghe. In Vladimir P. Nedjalkov, ed., *Reciprocal Constructions*, Vol. 2, 773–811. Amsterdam: John Benjamins.

Polinsky, Maria and Bernard Comrie. 2003. Constraints on Reflexivization in Tsez. In Dee Ann Holisky and Kevin Tuite, eds., *Current Trends in Caucasian, East European and Inner Asian Linguistics: Papers in Honor of Howard I. Aronson*, 265–289. Amsterdam: John Benjamins.

Schachter, Paul. 1977. Reference-Related and Role-Related Properties of Subjects. In Peter Cole and Jerrold M. Sadock, eds., *Syntax and Semantics 8: Grammatical Relations*, 279–306. New York: Academic Press.

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

van den Berg, Helma. 1995. *A Grammar of Hunzib (With Texts and Lexicon)*. Munich–Newcastle: Lincom Europa.

Bernard Comrie
Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6
D-04103 Leipzig, Germany

comrie@eva.mpg.de

Diana Forker
Universität Bamberg
Institut für allgemeine Sprachwissenschaft
Obere Karolinenstraße 8
D-96049 Bamberg, Germany

diana.forker@uni-bamberg.de

Zaira Khalilova
Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6
D-04103 Leipzig, Germany

khalilova@eva.mpg.de

The Unique Challenge of the Archi Paradigm[1]

GREVILLE G. CORBETT
University of Surrey

Introduction

The verbal paradigms of the Daghestanian language Archi are justly famous for their impressive size. I argue, however, that there is a more difficult problem lodged within a small and apparently simple part of the paradigm. It concerns the expression of gender and number, in their interaction with person. I present information on the large scale of the paradigm briefly, and then outline the problem of person (§1). The need, or lack of it, for a person feature in Archi has been discussed elsewhere, so that here I can simply summarize the argument (§2). If the need for a person feature is accepted, it follows that the paradigm has an unusual shape (§3). This paradigm is genuinely difficult, as I demonstrate in the main part of the paper (§4).

1 The Issue

The main reference on Archi is the four volume grammar and texts (Kibrik 1977a, b; Kibrik, Kodzasov, Olovjannikova and Samedov 1977a, b). A more accessible summary can be found in Kibrik (1998). Archi has a truly remarkable system of inflectional morphology. Kibrik (1998:466-468) calculates that a verb in principle has 1 502 839 forms. First there are tense/aspect/mood forms; if we add the

[1] I am grateful to the BLS organizers for the invitation to BLS, and to participants for fruitful discussion. Versions of the paper were given at "Ling Lunch" at the Université Paris-Diderot, UFR de Linguistique, 17 March 2011; the DATR Entropy Day, University of Brighton, 4 May 2011; the Max Planck Institute for Evolutionary Anthropology, Leipzig, 18 May 2011; and at the Workshop on Challenges of Complex Morphology to Morphological Theory, LSA Summer Institute, Boulder, Colorado, 27 July 2011. I thank these audiences for their suggestions, particularly Lynne Cahill and Farrell Ackerman. Special thanks are due to Marina Chumakina for sharing her expertise and insight; I am also grateful to Lisa Mack, Penny Everson and Claire Turner for help in preparing the materials. The support of the ERC (grant ERC-2008-AdG-230268 MORPHOLOGY) and of the ESRC (grant RES-062-23-0696) is gratefully acknowledged.

related gerunds, participles and masdars the total is already 12 405 (but note that some periphrastic forms are included here). If we add gender and number distinctions, and case-marked forms, the total rises to 188 463. The commentative, used for indirect speech, can be formed from all personal forms, and also from the admirative, and itself has an impressive array of forms; it is also the base for further participles. The additional forms (excluding gender and number distinctions) are 107 078. When gender/number and case distinctions are included that number rises to 1 314 376 forms. When added to 188 463 this gives 1 502 839 forms in total, as shown in (1):

(1) The number of forms derived from a single verb (Kibrik 1998:468)

plain/ commentative	without gender/number forms		with gender/number forms	
	without case forms	with case forms	without case forms	with case forms
plain	1 725	12 405	22 663	188 463
commentative	12 603	94 673	203 096	1 314 376
TOTALS	14 328	107 078	225 729	**1 502 839**

Impressive though these figures are, they are not our main concern. The system depends on a small number of basic stems, from which the large array of forms can be derived, as laid out in Chumakina (2011). These basic stems, with sound files, can be found in the Archi dictionary (Chumakina, Brown, Corbett and Quilliam 2007). Let us rather home in on the agreement markers, starting with the verb:

(2) Verbal affixes marking agreement in Archi

<table>
<tr><td rowspan="2">GENDER</td><td colspan="2">NUMBER</td></tr>
<tr><td>singular</td><td>plural</td></tr>
<tr><td>I (male human)</td><td>w-/<w></td><td rowspan="2">b-/</td></tr>
<tr><td>II (female human)</td><td>d-/<r></td></tr>
<tr><td>III (some animates, all insects, some inanimates)</td><td>b-/<b></td><td rowspan="2">Ø-/<Ø></td></tr>
<tr><td>IV (some animates, some inanimates, abstracts)</td><td>Ø-/<Ø></td></tr>
</table>

Archi has a complex agreement system, for which see Chumakina and Corbett

(2008). The main principle is that it has ergative-absolutive syntax, with agreement controlled by the absolutive argument. There are four genders and two numbers, with rather dramatic syncretisms, as (2) shows. The form before the slash is used in prefixal position, that after the slash is used in infixal position, with the important difference being in gender II singular. The bare stem is found for gender IV singular and for genders III and IV plural.

2 Person in Archi

Our next issue is whether (2) is sufficient, or whether verb agreement also needs to make reference to person. Archi has personal pronouns, distinguishing three persons, as well as clusivity, but it is not evident that person is a morphosyntactic feature. The issue has been discussed at length elsewhere, so here I give an outline, with references to more detailed discussion.

At first sight there is no evidence for a person feature (example (3) is from Marina Chumakina's fieldwork, (4) and (5) are from Kibrik, Kodzasov, Olovjannikova and Samedov (1977b:117, 121):

(3)	ɬːonnol	**d**-asːar-ši	**d**-i
	woman(II)[ABS.SG]	II.SG-tremble.IPFV-CVB	II.SG-be.PRS
	'The woman is trembling.'		

(4)	zon	**d**-irχːʷin
	1SG.ABS	II.SG-work
	'I work (woman speaking)	

(5)	un	hanžugur	**d**-aqˤa ?
	2SG.ABS	what.way	II.SG-come.PFV
	How did you get here? (to a woman)		

In each of these the verb, whether simple or periphrastic, agrees with the absolutive argument in gender and number (glossing follows the Leipzig Glossing Rules, for which see Comrie, Haspelmath and Bickel 2004, and II indicates the second gender). Consider now the plural pronouns (examples from Aleksandr Kibrik 1972 and personal communication):

(6)	teb	ba-qˤa
	3PL	I/II.PL-came
	'they (human) came'	

(7)	teb	qˤa
	3PL	[III/IV.PL]came
	'they (non-human) came'	

(8)	nen	qˤa	(9)	$\check{z}^{w}$en	qˤa
	1PL.EXCL	[?]came		2PL	[?]came
	‘we came’			‘you came’	

In the third person plural ((6) and (7)), agreement seems again to be simply a matter of gender and number. However, the first and second persons take what appears to be the wrong form: rather than the expected human plural (comparable to (6)), they have the same form as (7). So far, we have this picture:

(10) Agreement with personal pronouns in Archi

zon ‘I’	→	gender/number agreement
un ‘you (SG)’	→	gender/number agreement
teb ‘they’	→	gender/number agreement
nen / *nent’u* ‘we’ (EXCL/INCL)	→	bare stem
$\check{z}^{w}$en ‘you (PL)’	→	bare stem

Although the singular pronouns gave no evidence for person, the picture changes when we look at examples involving conjoining and resolution of feature values (Kibrik 1977b:187):

(11)	zoːn-u	buwa-wu	qˤa
	1SG.ABS-and	mother(II)[SG.ABS]-and	[??]come.PFV
	‘I and mother came.’		

The first person singular pronoun, which was apparently unproblematic on its own, causes problems when conjoined (as does the second person pronoun). Two solutions have been proposed: treating the difficulties within gender (Aleksandr Kibrik’s proposal) or recognizing a morphosyntactic feature of person in Archi. In brief, the consequences are as follows:

Option 1 (Kibrik et al. 1977a:63-64, Kibrik 1977b:186-187):

- Archi has no person feature
- the personal pronouns *zon, un, nen,* and *$\check{z}^{w}$en* form a special *gender*
- for resolution rules (based only on gender and number), genders must be *ranked*, with the gender containing the pronouns ranked higher than other genders

Option 2 (Corbett 1991:127-128, 271-273, Chumakina, Kibort and Corbett 2007):

- Archi has a person feature

- the gender resolution rules are unremarkable
- person resolution is standard (except that there is only the rule that persons 1 and 2 take precedence over person 3).

Clearly both options have some merit. We will not go into the detail here (for which see Chumakina, Kibort and Corbett 2007). Corbett (2012, chapter 8) makes the additional suggestion that embeddings of the values of one morphosyntactic feature in another (as implied by the first option) should be excluded in principle. For the rest of the paper, we assume option 2, and work out its interesting consequences for the morphology of Archi.

3 Proposed Verbal Agreement Paradigm in Archi

If we accept option 2, the paradigm of the verb is unusual:

(12) Gender, number and person in the Archi verb (first attempt)

PERSON	NUMBER	
	singular	plural
1	gender/number	bare stem (person)
2	gender/number	bare stem (person)
3	gender/number	gender/number

There is evidence for person in the first and second persons plural. Elsewhere the specification is for gender and number. This is already surprising, and there are further unusual points, which we discuss in turn.

4 What Is Special about the Archi Agreement Paradigm?

The Archi agreement paradigm is indeed remarkable. It is not the huge paradigm indicated in (1), interesting though that is, but rather the problem outlined in (12). We draw out its interest step by step.

4.1 Looking across the Lexicon, Only Some Items Agree

When a paradigm is presented, we often assume that it applies to all possible lexical items, an expectation gained from familiar languages of Western Europe perhaps. Archi is not like that. On the one hand there are some unexpected agreement targets, including adverbs. On the other hand, only some items in each part of speech show agreement at all. (13) gives data on the number of items which have an agreement paradigm: it is derived from Chumakina, Brown, Corbett and Quilliam (2007) and was reported in Chumakina and Corbett (2008:188); the figures for adverbs have been updated following reanalysis of some items:

(13) Numbers of items showing an agreement paradigm in Archi

	total	agreeing	% agreeing
verbs	1248	399	32.0
adjectives	446	313	70.2
adverbs	392	21	5.4
postpositions	34	1	2.9
enclitic particles	4	1	(25.0)

The numbers are surprising. In most parts of speech, it is only a minority of items which inflect for agreement features. In part it is a matter of having a stem of the right phonological shape, but according to current knowledge we require lexical specification of the items which agree or do not agree in many instances. The part of speech where we find a majority of agreeing items is the adjective; however, adjectives have a somewhat different paradigm, and are not involved in the person problem which is our main concern.

4.2 The Pattern of Cells Where There Is Agreement Can Be Remarkable

It is not sufficient to say that a lexical item agrees or not. We may have to specify which part of its paradigm is sensitive to incoming feature requirements. This is most clearly seen in the personal pronouns, given in (14):

(14) Personal pronouns: partial paradigms (Chumakina and Corbett 2008)

	SG		PL		
	1 person	2 person	1 person		2 person
			EXCL	INCL	
ABS	zon	un	nen	nen-t'-u	žʷen
ERG	zari		nen	nen-a-w nen-a-r-u nen-a-b-u nen-t'-u etc	žʷen
GEN	w-is \ b-is d-is / b-is \ is is /	wit	ulu d-olo b-olo olo etc	la-w-u la-r-u la-b-u la-t'-u etc	wiš
DAT	w-ez \ b-ez d-ez / b-ez \ ez ez /	wa-s	w-el d-el b-el el etc	w-ela-w d-ela-r-u b-ela-b-u el-t'-u etc	wež
COMIT	za-ł:u	wa-ł:u	la-ł:u		žʷa-ł:u
SIMILAT	za-qˤdi	wa-qˤdi	la-qˤdi		žʷa-qˤdi
COMP	za-χur	wa-χur	la-χur		žʷa-χur
SUBST	za-kł'ena	wa-kł'ena	la-kł'ena		žʷa-kł'ena
SUPERESS	za-t	wa-t	la-t		žʷa-t
SUPERELAT	za-t:i-š	wa-t:i-š	la-t:i-š		žʷa-t:i-š
SUPERLAT	za-t:i-k	wa-t:i-k	la-t:i-k		žʷa-t:i-k
SUPERTERM	za-t:i-kəna	wa-t:i-kəna	la-t:i-kəna		žʷa-t:i-kəna
CONTELAT	za-ra-š	wa-ra-š	la-ra-š		žʷa-ra-š
CONTLAT	za-ra-k	wa-ra-k	la-ra-k		žʷa-ra-k
CONTALL	za-r-ši	wa-ra-ši	la-ra-ši		žʷa-ra-ši
CONTTERM	za-ra-kəna	wa-ra-kəna	la-ra-kəna		žʷa-ra-kəna

There are several further cases not included here. The key point is that in some cells of the paradigm there is agreement (and according to the *pattern* in (2) though not all the forms match the verbal forms); however, this must be lexically specified, item by item and cell by cell. We shall see an example of this agreement in (22) below. For comparison with the situation in other languages of the family see Kibrik and Kodzasov (1990:220-223).

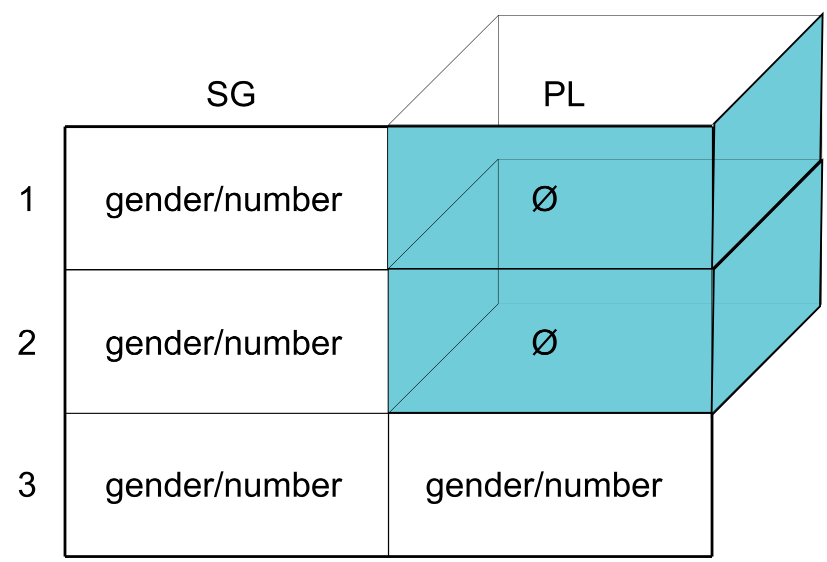
4.3 Where Agreement Is Possible, Almost All Items Show the Same Pattern of Gender/Number versus Person/Number

The surprising pattern summarized in (12) is not restricted to verbs. The pattern, though not the forms, are found more generally; the issue with person arises with the different agreement targets in (13); the exception, as already mentioned, is the adjective, which has a single form throughout the plural.

4.4 The Pattern of Person-Number versus Number-Gender Is Odd

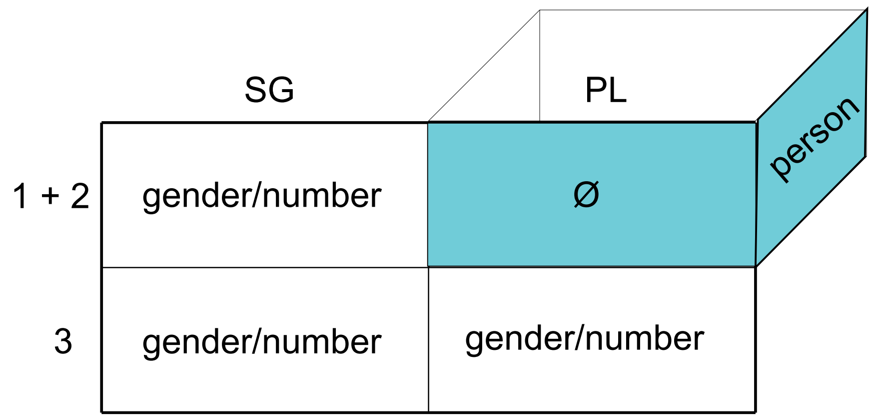
When we draw a paradigm in two dimensions, say for gender and number, we imply that we may need to make reference to either of the orthogonal features independently. If we have a third feature, this should ideally have its own dimension. It will be helpful to think in those terms here, and attempt to represent the Archi paradigm with an appropriate number of dimensions. The diagram in (15) is a first attempt:

(15) The dimensions of the Archi paradigm



This representation is partly right, in that it makes the point that there is a third dimension involved: the first and second persons plural are special in some way. However, the place of person is not fully clear (we return to this point in §4.6). Even so, (15) suggest that we can collapse the first and second persons, since the morphosyntax never distinguishes them. That gives us a simpler representation:

(16) The Archi paradigm simplified



(16) is a better representation, but still does not capture the place of person adequately (see §4.6 below).

4.5 The Slots with the Extra Dimension Have a *Morphomic* Pattern

When some cells in a paradigm show different behaviour from the rest, the pattern may be externally justified: the split may be justified in terms of grammatical meaning (for instance, we might find that all plural cells behave differently from all singular cells), or it may be justified in terms of phonology (for instance, all cells where the stem ends in a vowel behave differently from those where it ends in a consonant). But there are also examples where the split is purely morphology-internal, or 'morphomic'. Consider this partial paradigm from French:

(17) French *aller* 'go' in the present tense

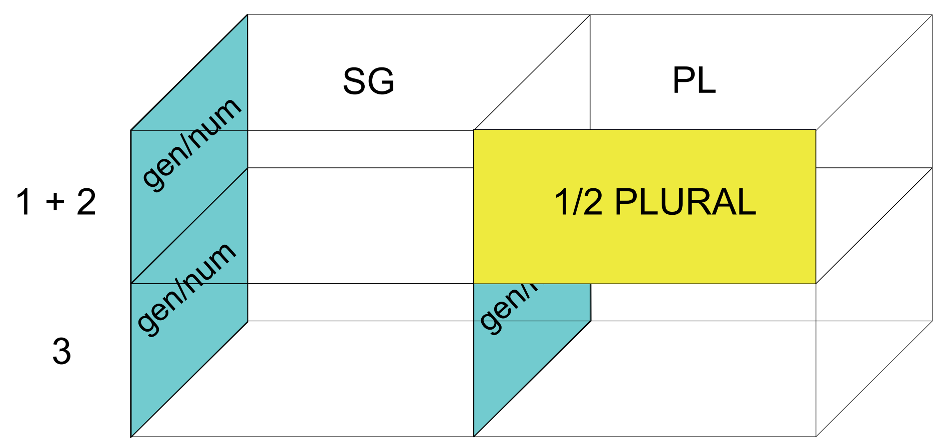
French 'go'	SG	PL
1	vais	allons
2	vas	allez
3	va	vont

Synchronically there is no external justification for the suppletive stem in the first and second persons plural of the present tense; the distribution is morphomic. Similarly, the Archi split in the paradigm is a morphomic one. The situation of the personal pronouns given in (14) is particularly interesting. The cells which show agreement have a morphomic distribution; and within each, the distribution of cells which agree in person is also morphomic. Thus we have a morphomic pattern nested under another morphomic pattern.

4.6 The Extra Dimension Defines a Small Part (the Tail Wags the Dog)

Let us return to the different dimensions of the paradigm. We noted that our earlier representation did not position person convincingly. Though person is available only in a small part of the paradigm, it determines the shape of the whole. Consider this alternative representation (18):

(18) Person as a determining feature in the paradigm



This representation clarifies the earlier problem. Person determines which other features are realized. It is a graphic case of the tail wagging the dog.

4.7 The Different Dimensions Are *Not* Based on Different Stems

When we find different behaviour in different parts of a paradigm, this can often be tied to different stems. Thus in Russian the past stem is sensitive to different features than the present stem. In the Archi paradigm, however, the dramatic difference we have seen based on person is not connected to any difference in the stem. The examples that we have seen, where there is no difference in stem for the section of the paradigm where person operates, are typical.

4.8 Person Is Marked by Syncretism Going across Word Classes

The syncretism we have observed between the 1/2 persons plural form and the gender III/IV plural form is not restricted to verbs. It is found with other agreement targets too. Significantly, it is found with those adverbs that agree, and here there is an overt marker. The paradigm in (19) is set out in the original gender/number format.

(19) An agreeing adverb in Archi: *dit:aw* ‘early, soon’ (gender/number forms)

GENDER	NUMBER	
	singular	plural
I	dit:a-**w**	dit:a-**b**-u
II	dit:a-**r**-u	
III	dit:a-**b**-u	dit:a-**t’**-u
IV	dit:a-**t’**-u	

The fact of having an overt marker should dispel any doubts about the reality of the syncretism (it is not a coincidence of uses of the bare stem, for instance). We can set out the same data including the person feature, as in (20):

(20) An agreeing adverb in Archi: *dit:aw* ‘early, soon’ (person/number)

PERSON	NUMBER	
	singular	plural
1/2	gender/number	dit:a-**t’**-u
3	gender/number	gender/number

4.9 The Syncretism Makes No Sense

Having established the syncretism, we now admit that it makes no sense. Two plural forms are available, basically human plural (genders I and II) and non-human plural (genders III and IV). Given that there is no dedicated form for person, it would seem evident which of these two forms would be chosen for agreement with first and second person pronouns in the plural. As we have seen, in Archi the other form is chosen systematically.

4.10 The Feature Person Is Non-Autonomous

In the simple instances, for each morphosyntactic feature we can point to some unique form; we justify postulating tense in English by pointing to forms like *computes* and *computed*. Similarly for each value: we justify having the value past by contrasting *computed* with other forms of the verb. There are instances, however, where values are proposed based on a combination of forms specified by values of another feature. Such instances are called 'non-autonomous' (Zaliznjak 1973:69-74). For example, in Classical Armenian, there is no dedicated form for the accusative (Baerman 2002). However, there is a pattern determined by number, which would lead us to postulate an accusative case. Transitive verbs govern forms which are as the nominative in the singular and as the locative in the plural. We may analyse these forms as accusative, and then the accusative case value is non-autonomous, since it has no dedicated form. However, what we are proposing for Archi is more extreme. In Armenian, and similar instances there is unique evidence for the feature case, it is only the particular value (accusative) which is non-autonomous. In Archi there is no dedicated form at all for the feature person: the feature, not a particular value, is non-autonomous.

4.11 Person Is Distinguished Only in Plural, So the Feature Is Not Realized Independently

Since morphosyntactic features are orthogonal to each other, they can be realized independently of each other. For instance, we may find an opposition of number for each case value, or indeed an opposition of person for each number value. In the Archi paradigm this is not so: we find evidence for person only in the plural. Thus person in Archi is not realized independently in this sense.

4.12 Inherent and Contextual Person and Number Values Can Conflict

We may draw a distinction between inherent and contextual features (Booij 1996, following Zwicky 1986). Inherent features are realized 'in the right place' (as when we find nominal number realized on nouns), while contextual features are there because of agreement or government. In the example *Victoria swims*, *Victoria* is inherently singular, while *swims* is contextually singular. Exceptionally an item may have both inherent and contextual features, and their values may conflict, as in this example from the Slavonic language, Upper Sorbian:

(21) Upper Sorbian (Faßke 1981:382-383; Corbett 1987)

moj-eho	muž-ow-a	sotr-a
my-M.SG.GEN	husband(M.SG)-POSS-F.SG.NOM	sister(F)-SG.NOM

'my husband's sister'

The head of the phrase, *sotra* 'sister' is inherently feminine, and it is also inherently singular. It is contextually nominative, as determined by its syntactic position. The adjective *mužowa*, derived from the noun *muž* 'husband', is also feminine, singular and nominative; these are all values of contextual features: the case value derives from the syntactic position, and the gender and number values are by agreement with the head noun *sotra* 'sister'. It is the possessive *mojeho* 'my' which shows the great interest of the construction. It is marked as masculine, singular and genitive. There is no expected agreement controller to account for these feature values: they do not match those of the head of the larger phrase, the noun *sotra* 'sister'. The only other candidate agreement controller is the possessive adjective *mužowa* 'husband's'. We know what the feature values of *mužowa* are, namely feminine and singular – which would not, of course, account for the form *mojeho*. One solution is to suggest that possessive adjectives of this type have both inherent and contextual features of number and gender, and that their values are independent of each other (see Stump 2001:15-17). In example (21), according to this solution, *mužowa* is contextually feminine and singular (through agreement with *sotra*), and is inherently masculine and singular, as for the noun *muž* 'husband'; it is these inherent values which *mojeho* agrees with).

The key point, then, is that features may be inherent or contextual, and that the same feature may be inherent and contextual on one and the same item; the values of the features are then independent of each other and may conflict. Having established this possibility, we return to Archi. We noted in (14) that some paradigm cells of the personal pronouns allow agreement, which is illustrated here (Kibrik 1994:349):

(22)	buwa-mu	**b**-ez	dit:a<b>u
	mother(II)-SG.ERG	SG.III-1SG.DAT	early<SG.III>
	χʷalli	a<b>u	
	bread(III)[SG.ABS]	made<SG.III>.PFV	
	'Mother made bread for me early.'		

The absolutive argument is *χʷalli* 'bread', and the verb *abu* 'made' agrees with it in gender and number. The adverb *dit:abu* 'early' is also in the gender III singular form to agree with *χʷalli* 'bread': see the paradigm in (19). Most interestingly, the first person singular pronoun in the dative case, *bez* 'to me', agrees: it too is gender III singular (see (14)). In Archi, the dative is also the case used with verbs of emotion and perception: thus in (23) the affected agent stands in the dative, and the object of perception takes the absolutive:

Archi (Bulbul Musaeva, thanks to Marina Chumakina)

(23)	ez	ž^{w}en	ak:u	dit:a<t'>u
	[1/2PL]1SG.DAT	2PL.ABS	[1/2PL]see.PFV	early<1/2PL>
	'I saw you (plural) early.'			

The object perceived is *ž^{w}en* 'you (plural)', in the absolutive case. The verb agrees with it, and has the bare stem. The adverb also agrees, and as we saw in (20) it has the infixed marker *t'*. Now consider the pronoun *ez*. From (22) we know that the first person singular pronoun, when in the dative, shows agreement. It is inherently first person singular, and its agreement is 1/2 plural (shown by the bare stem). In other words, its inherent and contextual features are in conflict. This is perhaps the most remarkable point about person in Archi. The inherent and contextual feature specification "had" to conflict in this way, yet it it seemed unimaginable, and it was therefore exciting to have the grammaticality of (23) confirmed.

5 Conclusions

What is special about the Archi agreement paradigm? The mammoth size of the verbal paradigm is of course remarkable. But the structure of the small part that involves person is perhaps of even greater interest. We noted twelve characteristics of the expression of person in the Archi paradigm. Each is of some interest. Their convergence on this small part of the paradigm makes it remarkable in the extreme.

References

Baerman, Matthew. 2002. Armenian Language Report. In: Surrey Syncretisms Database. Available at: http://www.smg.surrey.ac.uk/

Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In: Geert Booij and Jaap van Marle (eds) *Yearbook of Morphology 1995,* 1–15. Dordrecht: Kluwer.

Chumakina, Marina. 2011. Morphological complexity of Archi verbs. In: Gilles Authier and Timur Maisak (eds) *Tense, Aspect, Modality and Finiteness in East Caucasian Languages* (Diversitas Linguarum 30). Bochum: Brockmeyer.

Chumakina, Marina, Dunstan Brown, Greville G. Corbett and Harley Quilliam. 2007. *Archi: A Dictionary of the Archi villages, Southern Daghestan, Caucasus*. Available at: http://www.smg.surrey.ac.uk/archi/linguists/

Chumakina, Marina and Greville G. Corbett. 2008. Archi: the challenge of an extreme agreement system. In: A. V. Arxipov, L. M. Zaxarov, A. A. Kibrik, A. E. Kibrik, I. M. Kobozeva, O. F. Krivnova, E. A. Ljutikova and O. V. Fëdorova (eds). *Fonetika i nefonetika: K 70-letiju Sandro V. Kodzasova,* 184-194. Moscow: Jazyki slavjanskix kul´tur.

Chumakina, Marina, Anna Kibort and Greville G. Corbett. 2007. Determining a language's feature inventory: person in Archi. In: Peter K. Austin and Andrew Simpson (eds) *Endangered Languages* (special issue of *Linguistische Berichte,* number 14), 143-172. Hamburg: Helmut Buske.

Comrie, Bernard, Martin Haspelmath and Balthasar Bickel. 2004. The Leipzig Glossing Rules. Available at: http://www.eva.mpg.de/lingua/resources/glossing-rules.php. [Revised 2008.]

Corbett, Greville G. 1987. The morphology/syntax interface: evidence from possessive adjectives in Slavonic. *Language* 63:299-345

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press.

Faßke, Helmut. 1981. *Grammatik der obersorbischen Schriftsprache der Gegenwart: Morphologie*. Bautzen: Domowina Verlag.

Kibrik, A. E. 1972. O formal´nom vydelenii soglasovatel´nyx klassov v arčinskom jazyke. *Voprosy jazykoznanija* no. 1:124-31.

Kibrik, A. E. 1977a. *Opyt strukturnogo opisanija arčinskogo jazyka,* Vol. II: *Taksonomičeskaja grammatika*. (Publikacii otdelenija strukturnoj i prikladnoj lingvistiki 12). Moscow: Izdatel´stvo Moskovskogo universiteta.

Kibrik, A. E. 1977b. *Opyt strukturnogo opisanija arčinskogo jazyka*, Vol. III: *Dinamičeskaja grammatika*. (Publikacii otdelenija strukturnoj i prikladnoj lingvistiki, 13). Moscow: Izdatel´stvo Moskovskogo universiteta.

Kibrik, Aleksandr E. 1994. Archi. In: Rieks Smeets (ed.) *Indigenous Languages of the Caucasus IV: North East Caucasian Languages II: presenting The Three Nakh Languages and Six Minor Lezgian Languages,* 297-365. Delmar, NY: Caravan Books.

Kibrik, Aleksandr E. 1998. Archi. In: Andrew Spencer and Arnold M. Zwicky (eds) *The Handbook of Morphology,* 455-476. Oxford: Blackwell Publishers.

Kibrik, A. E. and S. V. Kodzasov. 1990. *Sopostavitel´noe izučenie dagestanskix jazykov: Imja: Fonetika*. Moscow: MGU.

Kibrik, A. E., S. V. Kodzasov, I. P. Olovjannikova and D. S. Samedov. 1977a. *Opyt strukturnogo opisanija arčinskogo jazyka*, Vol. I: *Leksika, fonetika*. (Publikacii otdelenija strukturnoj i prikladnoj lingvistiki, 11). Moscow: Izdatel´stvo Moskovskogo universiteta.

Kibrik, A. E., S. V. Kodzasov, I. P. Olovjannikova and D. S. Samedov. 1977b. *Arčinskij jazyk. Teksty i slovari*. (Publikacii otdelenija strukturnoj i prikladnoj lingvistiki, 14). Moscow: Izdatel´stvo Moskovskogo universiteta.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Zaliznjak, Andrej A. 1973. O ponimanii termina 'padež' v lingvističeskix opisanijax. In: Andrej A. Zaliznjak (ed.) *Problemy grammatičeskogo modelirovanija*, 53-87. Moscow: Nauka.

Zwicky, Arnold M. 1986. Imposed versus inherent feature specifications, and other multiple feature markings. In: *The Indiana University Linguistics Club 20th Anniversary Volume,* 85-106. Bloomington: Indiana University Linguistics Club.

Greville G. Corbett
Surrey Morphology Group
School of English and Languages
Faculty of Arts and Human Sciences
University of Surrey
Guildford, UK

g.corbett@surrey.ac.uk

Causativization and Contact in Nakh-Daghestanian

JOHANNA NICHOLS
University of California, Berkeley

1 Introduction

In the right sociolinguistic contexts, contact has been found to have predictable typological effects: inter-ethnic languages are less complex while isolated, local, non-interethnic languages are more complex (Szmrecsanyi 2009, Szmrecsanyi & Kortmann 2009, Trudgill 2009, 2011, Ross 1996, Nichols in press). This is because inter-ethnic languages are often simplified in the process of absorbing appreciable numbers of adult second-language learners, while no sociolinguistic force reduces the complexity of non-interethnic languages. Here I report a somewhat different example of such a correlation from the domain of verbal derivation, which has not figured prominently (if at all) in the literature on complexity. The languages surveyed come from the Nakh-Daghestanian (or East Caucasian) family of the eastern Caucasus (Russia, Georgia, Azerbaijan), an excellent natural laboratory for tracking the effects of social context on language structure. This is a very old language family with some 30 named languages and at least 35 actual languages (some of which are called "dialects" but are mutually unintelligible). (1) shows a schematic family tree.

(1) Nakh-Daghestanian subgrouping. * = language with mutually unintelligible dialects.

N-D	Nakh		Chechen*, Ingush, Batsbi
	Daghestanian		
		Avar-Andic	Avar
			Andic: Andi, Karata, Godoberi, Botlikh, Bagwalal, Chamalal, Tindi, Akhvakh*
		Tsezic	Tsez, Khwarshi, Hinuq, Hunzib, Bezhta
		Lak	Lak

Dargwa	Standard Dargi; Akusha, Uraxi, Kubachi, Chiragh, Mehweb, Ic'ari-Sanzhi, Xaidaq'
Lezgian	Lezgi, Tabassaran*, Aghul
	Rutul, Tsakhur
	Qrydz, Budukh
	Archi
	Udi
Xinalug	Xinalug (possibly a divergent branch of Lezgian)

2 Data and Survey

I survey three general typological properties across all the Nakh-Daghestanian daughter languages for which I could get the relevant data: complexity, transparency, and the number of verb pairs from a fixed list that are derived by overt causativization. Complexity is defined as the number of elements in a system, for a number of different subsystems from across the grammar. The elements and subsystems surveyed here are shown in (2). They are taken from Nichols 2009, with some additional structural variables specific to Nakh-Daghestanian.

(2) Complexity measures. Nakh-Daghestanian measures used here but not in Nichols 2009 are italicized.

Phonology: consonant series, vowels, tones, *phonation types*; syllable complexity

Classification: genders, possessive classes, *noun declension classes*

Inflectional synthesis of the verb: number of inflectional categories (following Bickel & Nichols 2005)

Syntax: alignments, basic word orders

Lexicon: inclusive/exclusive pronouns, *preverb slots, suppletive stems in first and second person pronouns*

The complexity values for the Nakh-Daghestanian languages measured on this scale range from 25 for Lezgi to 51 for Ingush.

Transparency is the extent to which each category has its own discrete marking, in other words the extent to which form mirrors meaning or categories. Kinds of non-transparency include suppletion, allomorphy, and semantically unpredictable categorization. (Transparency in several respects resembles canonical morphology as described for inflectional paradigms by Corbett 2007.) The transparency properties counted here are shown in (3). Transparency is more laborious to determine than complexity, and so far I have surveyed it for only two areas of the grammar.

(3) Transparency measures

Gender: allomorphy of gender markers; allopositionality of gender markers (e.g. prefixal only, prefixal in some verbs but infixal in others, etc.); predictability of gender from semantics

Argument coding: conjunct/disjunct agreement, hierarchical aligment

The gender measures have to do with how transparently the gender of a noun can be determined from its agreement marking and predicted from its semantics. An example of very transparent allomorphy is Avar or Chechen, where all gender agreement markers have a single allomorph, and there are no zero markers and no syncretism. These same languages have no allopositionality of gender markers: the markers on the verb are always root-initial. Less transparent positionality is common among Lezgian languages (e.g. Tsakhur: Dobrushina 1999), where some verbs take prefixal gender agreement and some infixal.[1] An example of transparent gender semantics is Avar, where in the singular all human males take *w-* agreement, human females take *j-*, and all other nouns take *b-*. Another is Tabassaran (Lezgian), where human nouns take *w-* and all others take *r-*. The maximum in semantic transparency of gender is found in the four languages that have lost gender entirely: Lezgi, southern Tabassaran, Aghul, and Udi. A gender system that is partly non-transparent is that of Ingush, shown in (4), where the gender of personal pronouns and human nouns is entirely predictable from person and the sex of the referent, while for other nouns gender classification is arbitrary.

(4) Ingush genders. *v, j, d, b* are the agreement prefixes. Their singular/plural pairings define as many as 8 genders.

	Sing.	*Plural*	*Examples*
1st, 2nd person pronouns	v/j	d	me, you, us
3rd person pronouns (human)	v/j	b	him, her, them
male human nouns	v	b	man, Ahmed (name)
female human nouns	j	b	woman, Easet (name)
some animals, inanimates	b	d	ox, head
some plants, inanimates	b	b	apple, family
inanimates, some animals	j	j	wolf, fence
inanimates, some animals	d	d	dog, house

The argument coding transparency measures have to do with how well one can predict the argument coding from the syntactic categories (or recover the syntactic categories from the coding). Conjunct/disjunct agreement patterns mark person, but mark it non-straightforwardly, wih the same form indicating first

[1] Infixal gender agreement results from entrapment of an agreement prefix when another prefix is added. In Lezgian languages the infixing verbs are generally old bipartite stems that are now largely fused and non-transparent synchronically.

person in questions and second person in statements (or vice versa), so that the relationship between person and marking is not straightforward. Hierarchical alignment marks syntactic relations, but causes argument roles to be obscured or marked indirectly because referential hierarchies (such as animacy or person) determine access to agreement slots.

The verb pairs surveyed are those of Nichols, Peterson, and Barnes 2004, who surveyed a fixed list of verb pairs across 80 languages and typologized languages by their derivational preferences. The verb pairs consist of a verb gloss and its semantic causative, e.g. 'fear' and 'scare, frighten', 'learn' and 'teach', etc. The set of 9 verb pairs surveyed is shown in (5).[2] The possible types of derivational treatment are listed in (6), and the first two are illustrated in (7).

(5) Verb pairs

Plain	*Semantic causative*
laugh	make laugh, amuse
die	kill
sit	seat, have sit, put in sitting position
eat	feed
learn	teach
see	show
be/get angry	anger, make angry
fear	scare, frighten
hide	hide

(6) Kinds of formal derivational treatment of the verb pairs in (5)

Causativization	(semantic causative is overtly derived)
Decausativization	(the plain verb is overtly derived)
Double derived	(both are derived)
Ambitransitive	(neither is derived, as with English *break*)
Ablaut or similar alternation	
Change in conjugation class only	
Change of light verb	
Plain verb is adjective, semantic causative is deadjectival verb	
Suppletion	

[2] Nichols, Peterson, and Barnes actually surveyed 18 verb pairs, 9 with prototypically animate S/O (e.g. 'fear' : 'scare') and 9 with prototypically inanimate S/O (e.g. 'break', intransitive and transitive). The 9 inanimate ones showed less typological variation and more sensitivity to universals than the others, and since they are all typically of lower text frequency than the animate ones they were less often to be found in dictionaries. These problems obtain for the Nakh-Daghestanian languages as well, so only the first 9, those with animate S/O, are used here.

(7) Examples of causativization (four Nakh-Daghestanian languages, above) and decausativization (two Slavic languages, below). Relevant derivational morphemes bold. Raised “c” = pharyngealization. “lh” = voiceless lateral fricative.

	‘fear’	*‘scare, frighten’*
Ingush (Nakh)	qier-	qiera-**d.ar**
Hunzib (Tsezic)	hi^{n}ch'a	hi^{n}ch'-**ek'**-a
Avar	h^{c}inq'ize	h^{c}inq'iz-**abize**
Godoberi (Andic)	lhibi	lhib-**al**-i
Macedonian	plaši **se**	plaši
Russian	bojat'-**sja**	pugat'

The languages surveyed are shown in (8). Not all of the Nakh-Daghestanian languages could be surveyed because the grammar survey requires fairly comprehensive descriptions and the survey of verb pairs requires a fairly comprehensive dictionary with an index. Coverage within branches is reasonably good but not optimal. (The Dargwa branch in particular is under-represented.)

(8) Languages surveyed. * = surveyed for verb pairs as well as complexity and transparency (others surveyed for complexity and transparency only).

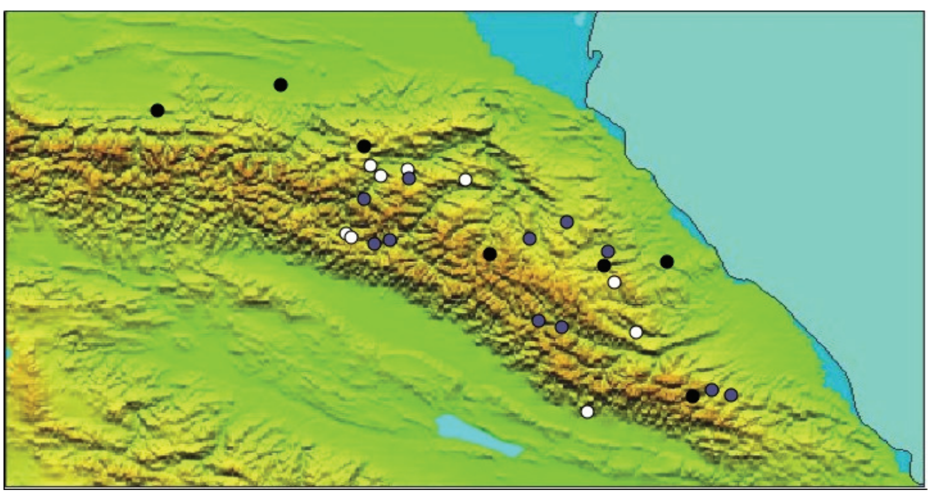
Nakh:	Ingush*
	Standard (lowland) Chechen*
	highland Chechen
Avar:	Standard Avar*
	Antsukh (southern) Avar
Andic:	Northern Akhvakh*
	Karata*
	Godoberi*
	Bagwalal*
Tsezic:	Tsez*
	Khwarshi
	Hinuq*
	Hunzib*
	Bezhta
Lak:	Lak*
Dargwa:	Standard Dargi*
	Kubachi
	Ic'ari
Lezgian:	Lezgi (standard)*
	Tabassaran (northern)*
	Aghul (Bursshag dialect)

Rutul*
Tsakhur*
Qrydz
Budukh*
Archi*
Udi*
Xinalug*

3 Findings

(9)-(11) plot levels of complexity, transparency, and causativization respectively, using the uniform convention that black dots = high values (high complexity, high transparency, high number of verb pairs with causativization), gray = medium, and white = low. The base map shows the eastern half of the Caucasus. Dots are placed at the main town where the language is spoken (for many of the highland languages there is only one town). For languages with large territories (chiefly Ingush, Chechen, Avar, Lezgi) the dot is placed at a historically important or central town (Ongusht for Ingush, Urus-Martan for lowland Chechen, Khunzakh for Avar).

(9) Complexity. White = low complexity, gray = medium, black = high.



For complexity (9), there is no particular distribution to high and medium levels (black and gray dots are scattered throughout the range), but there is a detectable pattern to low complexity (white dots), which forms clusters in two areas. To the west is a cluster of Avar-Andic and Tsezic languages (Avar slightly to the right, the Andic languages Karata, Bagwalal, and Godoberi to its left and at higher altitudes, and the Tsezic languages Tsez and Hinuq in the southern highlands). In geographical terms this means that Avar and nearly all of the Andic and

Tsezic languages along the Andi Koisu river have low complexity. (The remaining Andic language, Akhvakh, is just barely above the cutoff between low and moderate.) To the southeast, three Lezgian languages have low complexity: Aghul to the north, standard Lezgi, and Udi in the southern lowlands.

These two clusters coincide fairly well with the status of (current or past) inter-ethnic language. In the Avar-Andic-Tsezic area we see the results of long-term dominance of the Avar Koisu and Andi Koisu drainages by what is known as the Sarir kingdom (since the early first millennium BCE) and then the Avar Khanate (from its conversion to Islam until the Russian conquest of the eastern Caucasus in 1859). From the capital at Khunzakh (the location of the Avar dot on the map) it dominated the area economically, politically, and linguistically. As was typical throughout the Caucasus, highlanders needed to know lowland languages because the markets and winter pastures were there, while lowlanders did not need to know highland languages. Most of the working-age male highland population was transhumant, spending winters in the lowlands for winter pastures and seasonal work, and spending summers with their herds in highland pastures, and as a result most men were bilingual in their highland language and the lowland language. (See Volkova 1967, Wixman 1980, Nichols 2005.) Consequently, highland villages would occasionally shift to a lowland language, and lowland isoglosses and languages tended to move uphill. Now, the Avar-Andic subgroup is (impressionistically judged) of approximately Germanic-like diversity, and the Tsezic group, whose relatedness to Avar-Andic is widely but not universally accepted, is more divergent.[3] I assume that the stability of the Sarir/Avar dominance meant that there was a long-term spread of language from the Avar lowlands: first Proto-Avar-Andic-Tsezic, then Proto-Avar-Andic, then ancestral Avar spread along similar trajectories, so that the earliest branch to spread and diverge, Tsezic, is now in the highest highlands, the next branch, Andic, is in the lower highlands, and Avar has dominates the foothills, lowlands, and main river canyons. Human habitation of the Daghestanian highlands goes back millennia earlier, so each uphill spread replaced previously present languages by language shift. Thus at all times in the process the language spreading uphill was an inter-ethnic language much as Avar has been in historical times. The low complexity of most languages in the Avar sphere is consistent with their having been inter-ethnic languages. In the case of the Andic and eastern Tsezic languages the low complexity survives although the languages have not had inter-ethnic status for centuries or even millennia.

To the southeast the picture is probably similar. Lezgi and Aghul are very closely related.[4] Lezgi is a large language with a large speech community and large range, and is an inter-ethnic language in market towns in the nearby up-

[3] Korjakov 2006:21, 28 dates Avar-Andic to about 3500 BCE and Avar-Andic-Tsezic to slightly earlier, based on glottochronological counts that I have not reviewed.

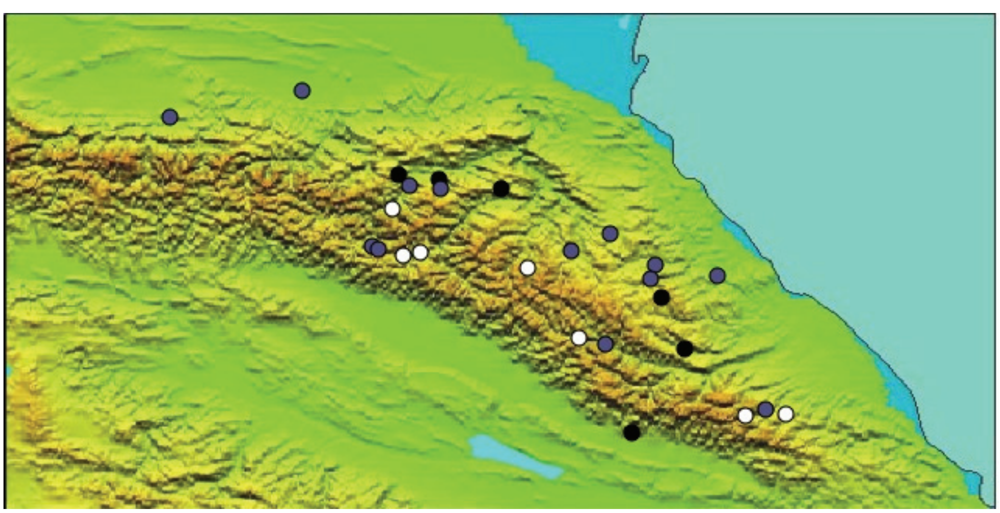
[4] Korjakov 2006:21 dates this branch at about 2500 years old, though see again note 3.

lands. Less is known about the ethnohistory of this area, but there is a long history of states and kingdoms dominating the Caspian coastal plain (which widens out in the Lezgi lowlands). Lezgi extends along the lower and middle Samur (the major river in the area) and its tributaries, and Aghul is adjacent and just above it on two major tributaries. This is a likely result of long-term spreading from the Samur lowlands, and the history of statehood means that there were important economic centers on the plain from which any major spread along these rivers must have emanated. Another close sister of Lezgi and Aghul, Tabassaran, is nearby as the crow flies but centered on a different river system and therefore probably not a result of the same spread. Its complexity is high.

Udi, the southernmost language and the only one centered on the south slope of the Caucasus range, is best known to linguists for its endoclisis (Harris 2002). It is a small language, spoken in two towns in Azerbaijan and a recently formed outpost in Georgia. Though an endangered enclave language now, Udi in the mid first millennium CE was an important language of the south Caucasian lowlands, known to philology as Alwan or Caucasian Albanian. It had a script created for it by Byzantine Christian missionaries, was an inscriptional language, and has a gospel translation in a recently discovered and published palimpsest (Gippert et al. 2009). Its low complexity is consistent with this history and, if that is its explanation, has lasted over a millennium.

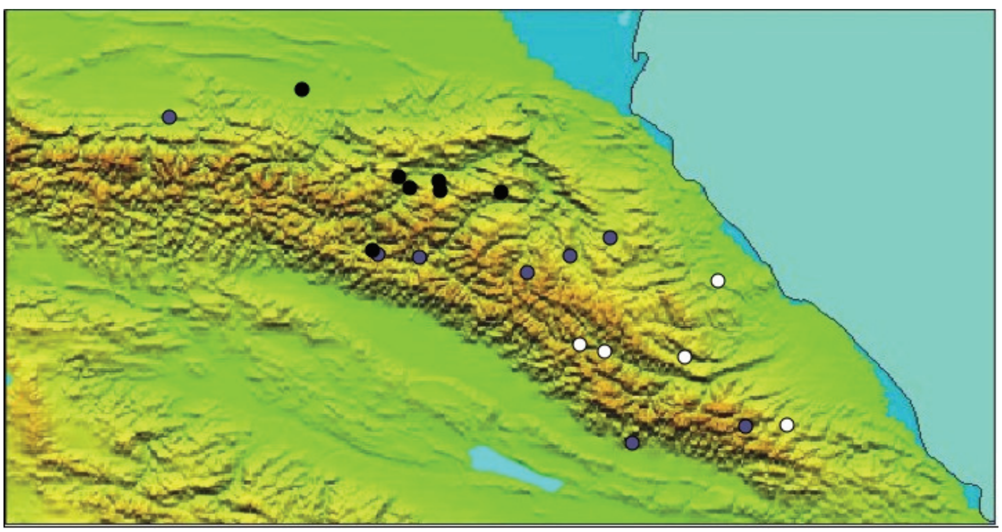
Transparency (10) gives a comparable picture. Low transparency (white dots) is found only in the highlands, and though the topography on the base map used here does not show this clearly most of the low-transparency languages are at the highest inhabited levels on their watercourses. These are languages that are sociolinguistically and geographically isolated and no known history of inter-ethnic use. High-transparency languages (black dots) again cluster in the Avar sphere and the Lezgian area, plus Udi.

(10) Transparency. White = low, gray = medium, black = high.



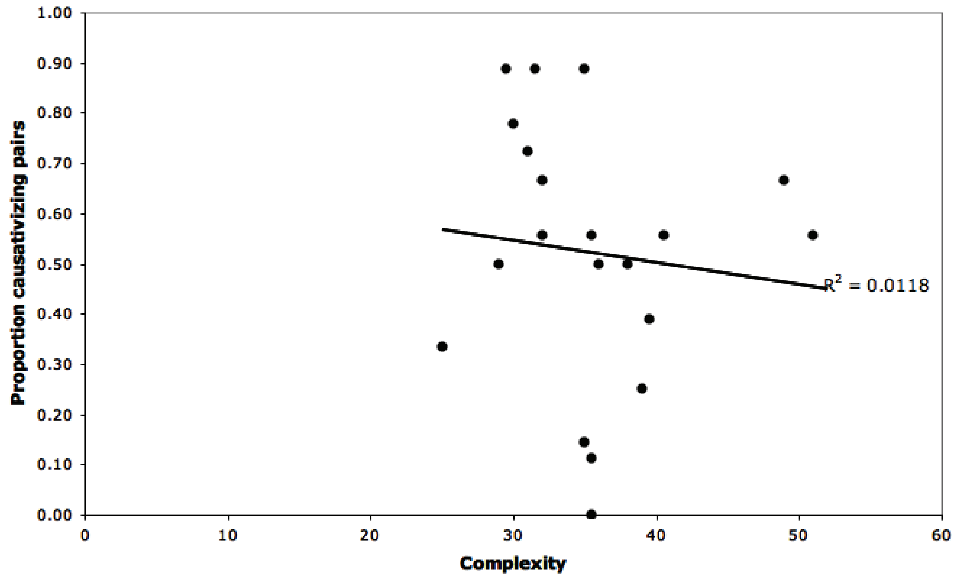
Causativization of verb pairs (11) shows a different distribution. Black dots are languages in which high numbers of the verb pairs are causativized. They form a large cluster in the Avar sphere, where Avar and some of the Andic languages have some of the world's highest proportions of causativized verbs (Creissels 2009), and they also include adjacent lowland Chechen. White dots have low numbers of causativized verbs. They include most of the Lezgian languages, all of which make extensive usage of light verb constructions and form many of their plain-causative pairs by using two different light verbs. Light verb constructions are frequent in the area comprising the southeast Caucasus, northern Azerbaijan, and nearby (Stilo 2009). Light verb constructions are a frequent typological correlate of the kind of lexicon in which simplex verbs are a closed class, and this too dominates in the same area. Most and perhaps all of the Nakh-Daghestanian languages have a closed class of simplex verbs and derive and borrow new verbs chiefly by forming light verb constructions. Therefore it is likely that the ancestral Nakh-Daghestanian type had a closed verb class and used an appreciable proportion of light verb constructions in its causative verb pairs, so that the preference for causativization in the Avar area is innovative.

(11) Proportion of the 9 verb pairs that are causativized

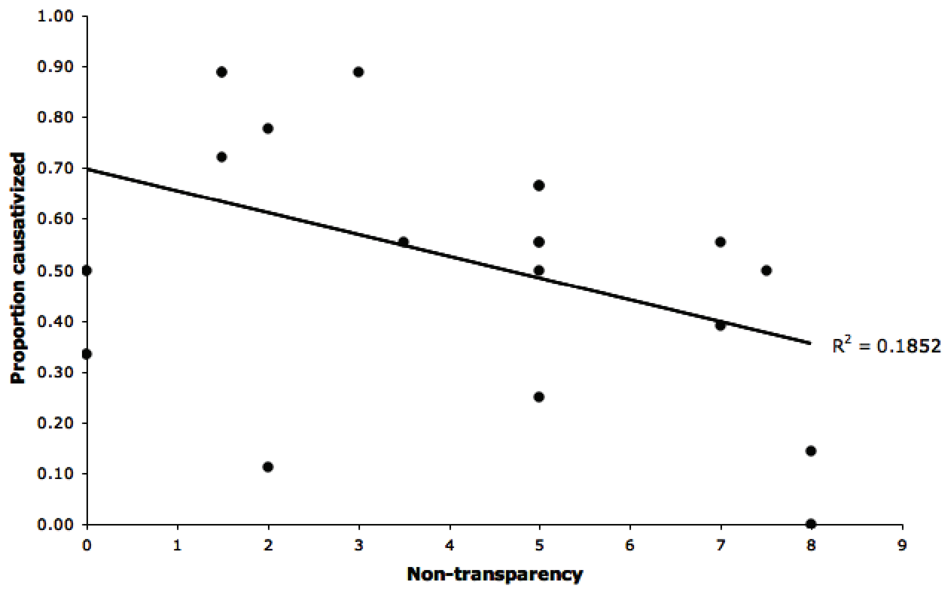


(12)-(13) show how complexity and transparency correlate with causativization, not in geographical clusters as just discussed but language by language. There is essentially no correlation with complexity; the trendline is nearly level. There is an appreciable correlation with transparency (measured as non-transparency in (12), so the negative correlation of causativization with non-transparency is a positive correlation of causativization with transparency). But even this correlation is not particularly strong, which suggests that it is not a purely typological correlation but rather is due to something in the historical contingencies of the two clusters of languages.

(12) Complexitivity and proportion of pairs causativized. (N = 20)



(13) Non-transparency and proportion of pairs causativized. (N = 20)



The causativizing type is itself an instance of high transparency. The usual analysis of pairs like ‘fear’ and ‘scare’ in syntax and semantics is that ‘scare’ consists of ‘fear’ plus causation. Assuming this reflects linguistic reality, then a morphological structure {fear-CAUS} is maximally transparent, and a lexicon that uses this structure widely is not only transparent but also non-complex in that the majority of verbs follow the same pattern. I suggest that this type has increased over time in the Avar sphere as repeated contact among the many small languages there, and contact of all of them with Avar, has made models of the transparent structure available and has favored borrowing and calquing of those models. Note that in the Avar area the smaller communities are politically autonomous and their languages are well retained; Avar is a lingua franca and there has been some shift to Avar but no whole-scale shift. An Avar variety was the language of command in the Avar army, but it was never imposed as a state or official language. The range of Avar as lingua franca extends well beyond its range as first language.

The sociolinguistics is different in the southeast, where causativization is not frequent. There, it is not that Lezgi serves as lingua franca among many small speech communities whose own languages remain autonomous; rather, if there were such languages, they have shifted to Lezgi, and the range of Lezgi as inter-ethnic language (one cannot really call it a lingua franca) does not extend far beyond its range as first language. The size of Lezgi is then due to an ordinary language spread, while the Avar-Andic-Tsezic region is an area of stable, long-term, and complex multilingualism. Both the Avar and the Lezgi situations have favored reduction of complexity and increase of transparency, but only the Avar one has fostered the long-term lexical influence that has favored spread of the most transparent model of verbal derivation, perhaps one word at a time (calque, loan translation, loan), gradually building up consistency across the lexicon.

There is some circumstantial evidence in favor of this analysis. There are three other areas in the world where strongly causativizing languages cluster: the eastern Eurasian steppe and nearby (Turkic, Mongolian, Tungusic, Tibetan); the Austronesian languages and some of their neighbors; and western North America (Nichols, Peterson, Barnes 2004, Fortescue 1998). These are all places where complex contact patterns involving many languages, probably including multilingualism and/or back-and-forth local language shifts, were the rule for long periods. If I am right about the Avar area, expansions of transparent patterns across the grammar and lexicon could have occurred in these places as well.

Thus the extralinguistic situation can plant seeds of change that interact with typological pressures to eventually produce an unexpected cluster of languages that are not true to their family's type.[5]

[5] Some of the work reported here was supported by NSF BCS 9222294, 9606448, and 0966675. Some of the research was done in the Max Planck Institute for Evolutionary Anthropology, Leipzig.

References

Bickel, Balthasar and Nichols, Johanna. 2005. Inflectional synthesis of the verb. In Martin Haspelmath, Matthew Dryer, Bernard Comrie, and David Gil, eds., *World Atlas of Language Structures*, 94-97. Oxford: Oxford University Press.

Corbett, Greville. 2007. Canonical typology, suppletion, and possible words. *Language* 83:1.8-42.

Creissels, Denis. 2010. P-lability and radical P-alignment. Presented at 19th International Symposium on Theoretical and Applied Linguistis, Workshop on Typology of Labile Verbs: Focus on Diachrony. Aristotle University of Thessaloniki, April 3-5, 2009.
Online: http://www.deniscreissels.fr/public/Creissels-P-lability.pdf
(accessed October 2011)

Dobrushina, Nina. 1999. Glagol. In A. E. Kibrik, ed., *Èlementy caxurskogo jazyka v tipologicheskom osveschenii*, 58-85. Moscow: MGU.

Fortescue, Michael. 1998. *Language Relations across Bering Strait*. London and New York: Cassell.

Gippert, Jost, Wolfgang Schulze, Zaza Aleksidze, and Jean-Pierre Mahé. 2009. *The Caucasian Albanian Palimpsests of Mount Sinai*. Turnhout, Belgium: Brepols.

Harris, Alice C. 2002. *Endoclisis and the Origins of Udi Morphosyntax*. Oxford: Oxford University Press.

Korjakov, Ju. B. 2006. *Atlas kavkazskix jazykov*. Moscow: Piligrim.

Nichols, Johanna. In press. The vertical archipelago: Adding the third dimension to linguistic geography. In Peter Auer, Martin Hilpert, Anja Stukenbrock, Benedikt Szmrecsanyi, eds., *Space in Language and Linguistics*, Berlin: Mouton de Gruyter.

Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil, and Peter Trudgill, eds., *Language Complexity as an Evolving Variable*, 110-125. Oxford: Oxford University Press.

Nichols, Johanna. 2005. The origin of the Chechen and Ingush: A study in alpine linguistic and ethnic geography. *Anthropological Linguistics* 46.129-155.

Nichols, Johanna, David A. Peterson, and Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8:2.149-211.

Ross, Malcolm D. 1996. Contact-induced change and the comparative method: Cases from Papua New Guinea. In Mark Durie and Malcolm D. Ross, eds., *The Comparative Method Reviewed*, New York: Oxford University Press.

Stilo, Don. 2009. An introduction to the Araxes-Iran linguistic area. MS, Max Planck Institute for Evolutionary Anthropology, Leipzig.

Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21:3.319-354.

Szmrecsanyi, Benedikt and Kortmann, Bernd. 2009. Between simplification and complexification: Non-standard varieties of English around the world. In Geoffrey Sampson, David Gil, and Peter Trudgill, eds., *Language Complexity as an Evolving Variable*, 65-79. Oxford: Oxford University Press.

Trudgill, Peter. 2009. Sociolinguistic typology and complexification. In Geoffrey Sampson, David Gil, and Peter Trudgill, eds., *Language Complexity as an Evolving Variable*, 98-109. Oxford: Oxford University Press.

Trudgill, Peter. 2011. *Sociolinguistic Typology: Social determinants of linguistic structure and complexity*. Oxford: Oxford University Press.

Volkova, N. G. 1967. Voprosy dvujazychija na Severnom Kavkaze. *Sovetskaja ètnografija* 1967:1.27-40.

Wixman, Ronald. 1980. *Language Aspects of Ethnic Patterns and Processes in the North Caucasus*. Chicago: University of Chicago.

Johanna Nichols
University of California, Berkeley
Department of Slavic Languages and Literatures, m.c. 2979
6303 Dwinelle Hall
Berkeley, CA 94720-2979, USA

johanna@berkeley.edu

Pharyngealization in Chechen is Gutturalization

JOHN SYLAK
University of California, Berkeley

Introduction

Knowing the phonetic and phonological properties of rare types of consonants, such as clicks, implosives, and pharyngeals, is essential for understanding how they affect the phonological systems of the languages in which they occur.[1] This study focuses on consonants with a primary or secondary pharyngeal articulation, which occur in only 21 of UPSID's 451 languages (5.32%; Maddieson 1984). However, these segments are found in over 12 different language stocks spread across North America, Eurasia, and Africa (Nichols and Bickel 2009). Pharyngeal or pharyngealized consonants, then, are rare enough token-wise that they are understudied in many respects, but are phylogenetically common enough that they are important to phonological theory and historical linguistics.

This study focuses on pharyngeal consonants and "pharyngealization" in Chechen, a Nakh-Daghestanian language of the northeast Caucasus region of the Russian Federation with approximately 1.3 million speakers (All-Russia Population Census 2002).[2,3] Previous accounts of pharyngeal consonants and "pharyngealization" in Chechen have, with one important exception, not included instrumental

[1] *Acknowledgements*: Many thanks to Johanna Nichols for inspiration, financial support, gathering field data, and being very patient. Thank you also to Sharon Inkelas, Keith Johnson, Andrew Garrett, and the audiences of the 2011 Berkeley QP Fest and BLS 37 for very helpful suggestions throughout this research. Thank you to UC Berkeley, the Beinecke Foundation, and the Survey of California and Other Indian Languages for additional financial support. Thank you, finally, to the Chechen speakers who contributed recordings. All remaining errors are solely my fault.

[2] This study seeks to explain data only from standard literary Plains Chechen, spoken in the lowlands surrounding Grozny. However, Chechen dialects offer pertinent material for a study on pharyngealization since it can be demonstrated that Plains Chechen has historically simplified $C^{ʕ}$ to C in some words (Magomedov 2005:125).

[3] I put the word *pharyngealization* in quotes when it is used as a cover term for what I will argue is both phonetic pharyngealization and epiglottalization. I do not use quotes when I mean purely phonetic pharyngealization as I define it in the discussion of acoustic tube modeling in §2.

data to support their claims, and moreover they do not agree on the basic phonemic inventory of Chechen. In trying to resolve that confusion, I have found evidence that what has been called "pharyngealization" in Chechen involves both phonetic pharyngealization and epiglottalization, and may be the acoustic result of a specific complex of muscle action that results in a flattened, backward protruding tongue configuration similar to that found in related neighboring languages such as Dargi (Gaprindašvili 1966), Tsakhur, and Udi (Ladefoged and Maddieson 1996:308; Catford 1983).

I first review the literature to highlight the diversity of opinions on how "pharyngealization" is to be analyzed and which pharyngeal consonants are present in the phonemic inventory of Chechen. Next, I explain the acoustic tube models that are used to generate predictions about the acoustic characteristics of phonetic uvularization, pharyngealization, and epiglottalization. I then compare these predictions to the output of linear regression models of acoustic data provided by 5 native speakers of Chechen. I bolster the conclusions drawn from those comparisons with evidence from Chechen phonology. I conclude by examining the broader implications of my findings for the idea of a GUTTURAL natural class and for understanding post-velar, supraglottal articulations.

1 Previous Accounts

Out of the literature that discusses "pharyngealization" in Chechen, only Kingston and Nichols (1987) provide instrumental phonetic data. Other works present a variety of opinions on which kinds of post-velar consonants should be considered part of Chechen's phonemic inventory, but no instrumental data to support their claims. To try to resolve this basic disagreement in the literature, this study attempts to determine the precise place(s) of articulation associated with what has been called "pharyngealization" in Chechen. After reviewing Kingston and Nichols (1987) in some detail, I present a short summary of the claims presented in the literature.

Kingston and Nichols (1987:15-18) examined "pharyngealization" in Chechen using recordings of wordlists provided by two native speakers (1 M, 1 F). The wordlists were composed of monosyllabic or disyllabic words which contained all the language's consonants and all the possible "pharyngealized" variants. All but two of the words contained /a/ or /aː/ in the first syllable, which always bears primary stress in Chechen. In most cases, the consonant of interest was word-initial. Kingston and Nichols (1987) extracted the frequencies of F1-F3 using an LPC technique, and found that "pharyngealization" involves a rise in F1 and a fall in F2 and/or F3, producing general compaction of the spectrum (Kingston and Nichols 1987:21). They also found that "pharyngealized" stops have a longer VOT than their plain or ejective counterparts (Kingston and Nichols 1987:18-19).

Nichols (1997:943,962-3) describes Chechen as having a pharyngeal stop (symbolized as /ʕ/) and fricative (/ħ/), and describes "pharyngealization" as "a morphophonemic prosody" associated with the preceding consonant. She argues that a cluster analysis (in which "pharyngealization" is analyzed as a consonant plus a following pharyngeal consonant) is disfavored because of "severe constraints on clusters, especially initial." She also argues against "pharyngealization" being associated with the vowel because its manifestation is centered between the consonant and vowel "in an almost segment-like acoustically compacted delay in voice onset of the vowel and in distortion of the formant transitions in the following vowel."

Aliroev (1999:42-44) analyzes Chechen as having two pharyngeal phonemes. One is a voiced pharyngeal stop with a voiced fricative allophone [ʕ] and the other is a voiceless fricative /ħ/. "Pharyngealization" is analyzed as a cluster of a consonant followed by an identically voiced pharyngeal phoneme.

Nichols and Vagapov (2004:21,35) present a phoneme inventory in which Chechen possesses an epiglottal stop and fricative. They describe "pharyngealization" as being phonetically epiglottalization.

Finally, Magomedov (2005:125) describes Chechen as having a single pharyngeal phoneme that varies between [ʕ] and [ħ]. He adopts a cluster analysis for pharyngealization.

2 Predictions from Acoustic Tube Modeling

Acoustic tube models can be used to make predictions about the formant values that will result from particular articulations (Stevens and House 1955; Fant 1960; Stevens 1998; Johnson 2003). One can simulate the effects of articulations at various places, including the uvula, middle pharyngeal wall, and epiglottis. The formant values generated by acoustic tube modeling can then be used as predictions of the properties that a sound made at a particular place of articulation will have. If the values from a given sound sample match the predicted values for a particular articulation, one has evidence that the articulation is being used. However, because the mapping from acoustics to articulation is 1:*many* (e.g. English /ɹ/ can be produced with two distinct articulations), a match between the predictions of acoustic tube modeling and the results of acoustic analysis can only be taken as evidence for, not proof of, the presence of the articulation coded into the model. This type of comparison is used by Shahin (2002) to analyze uvularization in Arabic and pharyngealization in St'át'imcets Salish and by Yeou (2001) and Yeou and Maeda (1995) to study pharyngeal consonants and "emphasis" in Arabic.

The acoustic tube models that are mathematically implemented here involve modeling the vocal tract as a combination of three tubes: one for the cavity formed behind the constriction, one for the constriction itself, and one for the cavity formed in front of the constriction. Three equations are used to describe the resonant frequencies (formants) that result as sound passes through these tubes. The equations describe general types of tubes, namely tubes closed at one or both ends or two tubes joined together as a resonant system (here, a Helmholtz resonator). The equation in (1) describes formants produced in the back cavity, which is modeled as a tube closed at both ends (Johnson 2003:106). The glottis forms one closed end since air flows only out of it (and into the tube). Because the constriction opening is small, little air escapes through it, and this opening can be considered to be effectively closed.

(1)

$$F_{bn} = \frac{nc}{2l_b}$$

n stands for the order of formant whose frequency is being calculated (first, second, third, etc.) and c stands for the speed of sound in the cavity, which is taken to be the speed of sound in warm, dry air (∼35,000 cm/sec). l_b stands for the length of the back cavity, which is determined by subtracting half the length of the constriction ($l_c/2$) from the point of constriction (measured as cm from the glottis).

The back cavity and the cavity formed by the constriction create a "resonant system called a Helmholtz resonator in which the volume of air in the constriction oscillates like a piston in and out of the constriction" (Johnson 2003:106). The single resonance produced by the Helmholtz resonator can be characterized by the equation in (2).

(2)

$$f = \frac{c}{2\pi}\sqrt{\frac{A_c}{A_b l_b l_c}}$$

A_c is the cross-sectional area of the constriction and l_c is the length of the constriction. A_b is the cross-sectional area of the back cavity while l_b is the length of that cavity.

The front cavity can be considered to be closed at one end and open at the other. The end of the cavity adjacent to the constriction can be considered closed because little air passes through. The other end is the opening formed by the lips. The resonances produced in the front cavity are described by the equation in (3), where l_f is the cavity's length (Johnson 2003:102).

(3)

$$F_{fn} = \frac{(2n-1)c}{4l_f}$$

l_f is calculated by subtracting the point of constriction and half the length of the constriction ($l_c/2$) from the total vocal tract length.

Vocal tract length was determined based on values measured from this study's data, but other parameters of the acoustic tube models were based on values reported in the literature on Arabic. Vocal tract measurements were made from recordings of the speakers by identifying (for each speaker) 5 points in time at which the first three formants were equally spaced. At these points, the vocal tract takes on a neutral shape, and the equation for a tube open at one end in (3) can be used to model the entire vocal tract (Johnson 2003:103). Because the formant frequencies are known in this instance, one can solve for the length variable of the equation. Doing so results in giving the equation the following form:

(4)

$$L = \frac{(2n-1)c}{4F_n}$$

I calculated a length for the vocal tract based on the F3 measurement at each of these 5 times. Those lengths were then averaged to provide a vocal tract length for each speaker. Finally, the lengths obtained for each speaker were averaged and the average vocal tract length for the speakers in the data was found to be 18.18 cm.

Using X-ray tracings from Ghazeli (1977) that were redrawn in Shahin (2002:31), the length of constriction was estimated for uvularized and pharyngealized articulations and was then scaled based on the average vocal tract length of 18.18 cm. The uvularized articulation was modeled with a length of constriction of 2.138 cm and the pharyngeal articulation with a length of constriction of 1.069 cm. The length of constriction for epiglottalized articulations was assumed to be 0.535 cm based on the size of the epiglottis in Ghazeli's (1977) X-ray tracings and how it makes contact with the pharyngeal wall.

The cross-sectional area of the various constrictions models the degree of constriction, and was determined with reference to the values used by Yeou and Maeda (1995), who obtained accurate predictions using certain values for voiced and voiceless uvular and pharyngeal consonants. For the uvular place of articulation, the value for the voiced uvular fricative and the value for the voiceless uvular fricative were averaged to obtain a cross-sectional area of 0.275 cm^2 because this study does not distinguish between voiced and voiceless consonants for the purpose of acoustic tube modeling or analysis. For the pharyngeal articulation, the voiced and voiceless fricatives were averaged to obtain the value 0.325 cm^2. The value for an epiglottal articulation was assumed to be 0.300 cm^2, which is the average of the values for the uvular and pharyngeal articulations as well as "the minimum cross-sectional area of the constriction for vowels which was measured by Fant (1960)" (Alwan 1986:28).

The cross-sectional area of the back cavity was assumed to be 2 cm^2 for a uvular secondary articulation. For a pharyngeal articulation, it was assumed to be 1.75 cm^2 since some sphincteric closure has been found to occur with pharyngeal and epiglottal articulations by Esling (1996:73-4). I assumed that such sphincteric closure would be greater with epiglottal articulations than with pharyngeal articulations, so the value 1.5 cm^2 was assumed for epiglottal articulations.

The points of constriction for uvular and epiglottal articulations were determined using X-ray tracings from Ghazeli (1977) in consultation with the parameters listed by Alwan (1986:28). When the points of constriction were scaled to the average 18.18 cm vocal tract used in these models, they were 8.019 cm from the glottis for uvular articulations and 3.742 cm from the glottis for epiglottal articulations. The point of constriction for a purely pharyngeal articulation was assumed to lie halfway between these two, and was assumed (after scaling) to be 5.881 cm from the glottis.

To account for raising of the larynx in pharyngeal and epiglottal articulations, which is reported by Esling (2005:21) and measured quantitatively by Alwan (1986:28), a small amount was subtracted from the back cavities associated with those two places. For the pharyngeal articulation, 0.3743 cm was subtracted, while for the epiglottal articulation, 0.7486 cm was subtracted. The subtraction for the epiglottal place of articulation is based on Alwan's (1986:28) measurement of 0.7 cm of larynx raising during Arabic pharyngeal consonants, which she notes are associated with backward and downward movement of the epiglottis. The subtraction for what this study calls the pharyngeal place of articulation is assumed to be 0.35 cm, half of the measured 0.7 cm, in Alwan's model. After scaling 0.35 cm and 0.7 cm to the vocal tract length used in this model (18.18 cm), the subtractions are 0.3743 cm and 0.7486 cm for the pharyngeal and epiglottal places of articulation.

The following table summarizes the parameters used in the acoustic tube models to derive the formant frequencies for the three possible secondary articulations. All the values for these parameters are in centimeters unless otherwise noted.

(5) Parameters Used in Acoustic Tube Models

	Uvularization	Pharyngealization	Epiglottalization
Vocal Tract Length	18.18	18.18	18.18
Point of Constriction	8.019	5.881	3.742
l_c	2.138	1.069	0.535
A_c (in cm^2)	0.275	0.325	0.300
A_b (in cm^2)	2.00	1.75	1.50
Adjustment to Back Cavity	—	-0.3743	-0.7486
Back Cavity Length[4]	6.950	4.972	2.726
Front Cavity Length	9.088	11.761	14.167

Using these parameters, nomograms were produced to derive formant frequencies from the acoustic tube models. These are not shown for reasons of space but are available upon request and can be seen in Sylak (2011). The nomograms varied the point of constriction but held all other values constant. The table below shows the predictions that are compared to the results of acoustic analysis in the next section.

[4] The adjustment for larynx raising is incorporated into these values.

(6) Summary of Predictions

All values in Hz	F1	F2	F3	F2−F1	F3−F1
Uvularization	536	963	2518	427	1982
Pharyngealization	744	1041	2232	297	1488
Epiglottalization	618	1853	2063	1235	1445

3 Acoustic Analysis

3.1 Data

The data for acoustic analysis come from recordings of readings of a wordlist by 5 male native speakers of Chechen originally from the Republic of Chechnya. The wordlist elicited all the "pharyngealized" versus plain consonant contrasts and elicited many of the possible vowels after a glottal stop (a plain consonant) and after an epiglottal stop (which can be thought of as a "pharyngealized" glottal stop; Kingston and Nichols 1987). Praat transcription (Boersma and Weenink 2001) was used to delineate the vowel after a pharyngealized consonant or its plain counterpart since the vowel, especially the first half, is where the effects of "pharyngealization" from a preceding consonant are greatest (Kingston and Nichols 1987; Yeou 2001). These delineated vowels, which were always /aː/ or /a/, were sorted based on the place of articulation of the consonant preceding them (labial, dental, alveolar, or post-alveolar). When the effects of "pharyngealization" on different vowel qualities was being examined, vowels were sorted into the groups front non-low, back non-low, and low because these seem to correspond to three main types of tongue configurations (Ladefoged and Maddieson 1996:284, fig. 9.3). For reference, the following table shows the phonemic inventory of Chechen, following Nichols and Vagapov (2004:21). The segments that were examined in this study are indicated in bold. Where I depart from the phonetic transcription of vowels below in the rest of the study, I indicate the variant that I use in parentheses introduced with an equal sign. This variant conforms to Nichols' and Vagapov's (2004) working romanization.

(7) Phonemic Inventory of Chechen (Nichols and Vagapov 2004:21)

CONS	Labial	Dental	Alveolar	Postalv	Velar	Uvular	Ep/Ph	Glot
Stop	**p** p' **b**	**t** t' **d**	**t͡s** t͡s'	**t͡ʃ** t͡ʃ'	k k' g	q q'	**ʡ**	**ʔ**
Fricative	v		**s** **z**	ʃ **ʒ**	x	ɣ	**ħ**	**h**
Nasal	**m**		**n**					
Liquid		l	r̥ r					
Glide				j (= pal.)				

VOWELS	Front	Front Round	Central/ Back	Back
High	**i iː** ɛ eː	**y** yː		**u uː** **o oː**
Mid	**^{i}ɛ/^{i}æ iːe̯** (= **ia ie**)	y̯œ yːo	ʌ/ə(= a)	u̯ɔ uːo̯/uːə̯
Low	**æ**		**a aː(= aa)**	o̯a/ɔ ɔː

3.2 Methods

Praat was used to separate and transcribe C, C_P, and V in contrasting C_PV and CV sequences (C_P = "pharyngealized" consonant). Out of C, C_P, and V, it was V that most often showed a statistically significant difference in its formants between its realizations in a CV and a C_PV sequence. Thus, V was chosen as the object of analysis. Praat's formant tracking and scripting capabilities were used to obtain 10 equally temporally separated measurements for F1-F3 in each V that was transcribed in the relevant C_PV and CV sequences. These measurements were taken in order to gather more accurate data on formant trajectories and were hand-corrected where necessary.

3.3 Analysis

To see how the formant trajectories differed between segments with respect to time, pharyngealization status, speaker, and vowel, linear regression modeling was used.[5] Because the formant measurements were taken over brief time periods (from approximately 55 to 300 milliseconds) for monophthongs, the data for each formant was assumed to be roughly linear. The statistical program R (R Development Core Team 2011) was used to computationally implement the linear regression models. Linear regression models were used to analyze the 10 equally temporally separated values of F1, F2, and F2−F1 according to the consonantal or vocalic subgroup assigned to the data being analyzed. The results of applying linear regression to the F3 data are not reported since the difference between a plain and pharyngealized consonant was almost never found to be significant. F1 and F2 have both been reported to be salient to the perception of "pharyngealized" consonants, as has the value of F2−F1 since it models compaction of the spectrum (Kingston and Nichols 1987). Compaction of the spectrum was identified as an analytically (and probably perceptually) noticeable effect of pharyngealization in Chechen (ibid.).

Because /aː/ or /a/ was the vowel most frequently encountered in data on the consonantal subgroups, vowel quality was not assumed to have a main effect in the

[5] Many thanks to Melinda Woodley for suggesting this method, for helping to implement it, and for advice on how to interpret the results.

models. However, speaker identity was assumed to have an effect since as few as 3 speakers may have provided data for a given "pharyngealized" vs. plain contrast. This could skew the data, and so the effect had to be taken into account. The linear regression model used to analyze the property of interest (F1, F2, F2−F1; abbreviated hereafter as 'PI') for consonantal subgroups was:

(8) PI∼Time*Pharyngealization+Speaker

It was assumed that F1, for example, would start out higher in pharyngealized segments and fall as time passed. Thus, it was assumed that time and pharyngealization would both be main effects and that they would interact, with a given formant measurement (e.g. F1) increasing or decreasing through time in pharyngealized variants but staying constant in plain variants.

For analyzing properties of interest according to vocalic subgroups, the linear regression model that was used is that in (9).

(9) PI∼Time*Pharyngealization+Speaker+Vowel

Because the starting values of the properties of interest were highly dependent on the quality of the vowel being examined, vowel quality was assumed to be a main effect.

3.4 Results

At this point, it is possible to compare the results of acoustic analysis to the predictions made via acoustic tube modeling. In the following discussion, I will proceed from anterior to posterior by place of articulation through the consonantal subgroups and then through the vocalic subgroups in the order front non-low, back non-low, low. For reasons of space, I do not report full results for the linear regression models.[6] Instead, I report the sum of each linear regression model's intercept plus the main effect of "pharyngealization" in order to obtain a value for each formant in each consonantal and vocalic subgroup. The intercept can be thought of as a baseline value that one might expect to occur after a plain consonant. The main effect of "pharyngealization" provides an estimate of how much one can expect the actual formant measurement to deviate from the intercept when the token is after a "pharyngealized" consonant.

Each table in (10)-(16) shows the value (intercept+pharyngealization) from the linear regression model under the label 'Measured,' followed by a reminder of the predictions made by acoustic tube modeling (shown originally in (6)) of the formant values each secondary articulation should yield. Finally, the table shows my judgment of which secondary articulation there is evidence for. Above each table, I show the place of articulation that was analyzed, the segments at that place that

[6] For full results and a longer version of this paper, see Sylak (2011).

were analyzed, and my overall judgement of which secondary articulation is occurring at that place. The table in (17) gives an overall summary of which secondary articulations may be occurring at each place of articulation.

(10) Labial (p, b, m): *Pharyngealization*

	Measured	Uv.	Ph.	Ep.	Secondary Articulation Matched
F1	772.652	536	744	618	Pharyngealization
F2	998.687	963	1041	1853	Uvularization/Pharyngealization
F2−F1	226.035	427	297	1235	Pharyngealization

(11) Dental (t, d): *Pharyngealization*

	Measured	Uv.	Ph.	Ep.	Secondary Articulation Matched
F1	746.493	536	744	618	Pharyngealization
F2	1103.455	963	1041	1853	Pharyngealization
F2−F1	356.961	427	297	1235	Pharyngealization

(12) Alveolar (ts, s, z, n): *Pharyngealization*

	Measured	Uv.	Ph.	Ep.	Secondary Articulation Matched
F1	775.962	536	744	618	Pharyngealization
F2	1182.646	963	1041	1853	Pharyngealization
F2−F1	406.684	427	297	1235	Uvularization

(13) Post-Alveolar (ʧ, ʒ): *Epiglottalization* or *Pharyngealization*

	Measured	Uv.	Ph.	Ep.	Secondary Articulation Matched
F1	638.777	536	744	618	Epiglottalization
F2	1144.726	963	1041	1853	Pharyngealization
F2−F1	505.948	427	297	1235	Uvularization

(14) Front, Non-Low vowels (i, iː, y, yː, ia, ie): *Epiglottalization*

	Measured	Uv.	Ph.	Ep.	Secondary Articulation Matched
F1	432.4	536	744	618	Uvularization
F2	1903.353	963	1041	1853	Epiglottalization
F2−F1	1470.953	427	297	1235	Epiglottalization

(15) Back, Non-Low vowels (u, uː, o, oː): *Epiglottalization* or *Pharyngealization*

	Measured	Uv.	Ph.	Ep.	Secondary Articulation Matched
F1	584.099	536	744	618	Epiglottalization
F2	1010.73	963	1041	1853	Pharyngealization
F2−F1	426.631	427	297	1235	Uvularization

(16) Low vowels (a, aː, æ): *Pharyngealization*

	Measured	Uv.	Ph.	Ep.	Secondary Articulation Matched
F1	810.327	536	744	618	Pharyngealization
F2	1144.648	963	1041	1853	Pharyngealization
F2−F1	334.321	427	297	1235	Pharyngealization

(17) Summary of Results of Acoustic Analysis

LABIAL	DENTAL	ALVEOLAR
Pharyngealization	Pharyngealization	Pharyngealization
POST-ALVEOLAR		
Epiglottalization *or* Pharyngealization		

FRONT NON-LOW	BACK NON-LOW	LOW
Epiglottalization	Epiglottalization *or* Pharyngealization	Pharyngealization

3.5 Discussion

What is called "pharyngealization" in Chechen seems actually to be *two* phonetic types of secondary articulation: pharyngealization and epiglottalization. A possible explanation for why these articulations are grouped into one effect ("pharyngealization") is that they are the results of a single complex of muscle action in and around the tongue. This complex of muscle action produces a tongue configuration that is affected by other muscle actions necessary for achieving the primary articulation of the segment in question. This is what causes "pharyngealization" to be realized variously as both pharyngealization and epiglottalization. Moreover, the complex of muscle action seems to produce consistent acoustic effects (such as elevated F1 and lowered F2; Kingston and Nichols 1987) that are perceived as belonging to a single phonological modification.[7]

One complex of muscle action that could produce what has been called "pharyngealization" in Chechen is the constriction of the inferior pharyngeal constrictor

[7] An alternative explanation may be that pharyngealization and epiglottalization produce similar enough acoustic effects that listeners perceive them as the same articulation.

bringing the tongue back and the vertical muscle flattening the tongue (Abd-El-Malek 1939; Ladefoged et al. 2002). This gives the tongue a flattened, plateau-like shape and causes it to protrude posteriorly toward the pharyngeal wall. A similar tongue configuration is directly attested by Gaprindašvili (1966:14) for Dargi and by Catford (1983) for Udi and Tsakhur, which are all related to Chechen.

Specifically, the complex of muscle action may operate in the following way. When the tongue tip must be brought forward for dental and alveolar articulations, the protrusion of the tongue backward may be hampered. This causes pharyngealization because the tongue protrudes backward, but at a higher point. In addition, because the tongue is flattened to some degree, the back protrusion does not produce uvularization, which would require raising and arching of the tongue. When the tip of the tongue must be raised significantly, as with a post-alveolar consonant, the tongue is still flattened and protruded backward, but epiglottalization is produced because the raising of the tongue tip causes the back protrusion to lower, like a seesaw. With vowels, the tongue cannot be as significantly flattened as it can be with anterior consonants. However, since the tongue's mass is shifted backward, the action of the inferior pharyngeal constrictor is more pronounced. This may be the explanation for why front and back non-low vowels are associated with epiglottalization. With low vowels, the constriction point for the vowel is so near the middle pharynx wall that the inferior pharyngeal constrictor is already constricting to that position, making it unavailable for constriction at another point.

4 Evidence from Chechen Phonology

While the mapping from articulation to acoustics is 1:1 (the same articulation will always yield the same acoustic result), the reverse mapping from acoustics to articulation is 1:*many*. This means that when the results of acoustic analysis match the predictions from acoustic tube models, there is *evidence* for the articulation embodied in the prediction from the acoustic tube model, but not definitive *proof* since another articulation might be able to produce the same acoustic effects. This means that additional evidence must be sought in the absence of articulatory data, and Chechen phonology provides that evidence.

One fact about Chechen phonology that can be explained by the proposed complex of muscle action is the fact that velar and uvular consonants cannot be "pharyngealized" (Nichols 1997:963). With a velar or uvular consonant, the tongue dorsum is forced up to the velum or uvula, but such an upward forcing of the tongue is antithetical to the flattening action that is part of the proposed complex of muscle action. Thus, "pharyngealization" cannot occur with these consonants.

Another fact about Chechen phonology that can be explained with this complex of muscle action is that "pharyngealization" is a free variant of a syllable-initial post-consonantal uvular ejective (Magomedov 2005:125) as in the words *t͡ʃq'or* [t͡ʃˤor] "bark, skin" and *tq'a* [tˤʌ] "twenty" (Nichols and Vagapov 2004:270,405). A possible explanation is that the first consonant positions the tongue for the primary articulation while the uvular stop positions the tongue dorsum in its uppermost and backmost position (at the uvula). If a speaker relaxes the uvular articulation, the tongue dorsum falls and is susceptible to being flattened by the relaxation of the transversal muscles, yet the tongue is still far back. In addition, the larynx raises in preparation for the ejective release of /q'/, which creates an acoustic effect similar to pharyngealization or epiglottalization, since these are accompanied by larynx raising (Esling 1996; Alwan 1986). This creates an effect similar enough to that created by the proposed complex of muscle action that speakers hear "pharyngealization," although in fact the tongue may not be actively flattened by the vertical muscle or pulled posteriorly by the constriction of the inferior pharyngeal constrictor. The proposal of a specific complex of muscle action seems to be supported both by predictions from acoustic tube modeling and by evidence from Chechen phonology.

5 Conclusion

By examining "pharyngealization" in detail, this study has shown that the pharyngeal and epiglottal places of articulation are not phonologically contrastive in Chechen, as opposed to what has been claimed for Agul (Ladefoged and Maddieson 1996:37-8). In addition, it has been shown that "pharyngealization" varies freely with a uvular articulation, /q'/, in the post-consonantal position of a syllabic onset. From these facts, one can infer that the phonetically uvular, pharyngeal, and epiglottal places of articulation are phonetically and phonologically grouped together, at least by the phenomena shown. I interpret this as evidence supporting the existence of a GUTTURAL natural class in Chechen (McCarthy 1994). If one chooses to interpret the evidence this way, as I do, then "pharyngealization" is better viewed as *gutturalization*, since "pharyngealization" has been shown here to involve both phonetic pharyngealization and epiglottalization.

References

Abd-El-Malek, Shafik. 1939. Observations on the Morphology of the Human Tongue. *Journal of Anatomy* 73:201–210.

Aliroev, Ibragim Junusovič. 1999. *Čečenskij Jazyk*. Moskva: Academia.

Alwan, Abeer Abdul-Hussain. 1986. Acoustic and Perceptual Correlates of Pharyngeal and Uvular Consonants. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Boersma, Paul and David Weenink. 2001. Praat, a System for Doing Phonetics by

Computer. *Glot International* 5(9/10):341–345. Includes review by Vincent van Heuven.

Catford, John Cunnison. 1983. Pharyngeal and Laryngeal Sounds in Caucasian Languages. In Diane M. Bless and James H. Abbs, eds., *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, 344–350, San Diego: College Hill Press.

Esling, John H. 1996. Pharyngeal Consonants and the Aryepiglottic Sphincter. *Journal of the International Phonetic Association* 26(2):65–88.

Esling, John H. 2005. There Are No Back Vowels: The Laryngeal Articulator Model. *The Canadian Journal of Linguistics* 50(1):13–44.

Fant, Gunnar. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

Gaprindašvili, S. 1966. *Fonetika Darginskogo Jazyka*. Tbilisi: Mecniereba.

Ghazeli, Salem. 1977. Back Consonants and Backing Coarticulation in Arabic. Ph.D. thesis, University of Texas at Austin, Austin, TX.

Johnson, Keith. 2003. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell, 2nd edition.

Kingston, John and Johanna Nichols. 1987. Pharyngealization in Chechen: Its Implications for Phonetics and Phonology. Ms. Earlier version presented at 112th meeting of the ASA in Anaheim, CA, Dec. 1986.

Ladefoged, Peter, Melissa Epstein, and Narineh Hacopian. 2002. Dissection Manual for Students of Speech. *UCLA Working Papers in Phonetics* 102.

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

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge, UK: Cambridge University Press.

Magomedov, Arslanbek Gafurovič. 2005. *Očerki Fonetiki Čečenskogo Jazyka*. Maxačkala: Dagestanskij Naučnyj Centr RAN.

McCarthy, John J. 1994. The Phonetics and Phonology of Semitic Pharyngeals. In Patricia A. Keating, ed., *Phonological Structure and Phonetic Form*, *Papers in Laboratory Phonology*, volume III, 191–233, Cambridge, UK: Cambridge University Press.

Nichols, Johanna. 1997. Chechen Phonology. In Alan S. Kaye and Peter T. Daniels, eds., *Phonologies of Asia and Africa (Including the Caucasus)*, Winona Lake, IN: Eisenbrauns.

Nichols, Johanna and Balthasar Bickel. 2009. The AUTOTYP Genealogy and Geography Database: 2009 Release. URL: `http://www.uni-leipzig.de/∼autotyp`.

Nichols, Johanna and Arbi Vagapov. 2004. *Chechen-English and English-Chechen Dictionary*. New York: RoutledgeCurzon.

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Freely available open source software distributed under GNU General Public License at `http://www.r-project.org/`.

Shahin, Kimary N. 2002. *Postvelar Harmony*. Amsterdam: John Benjamins.

Stevens, Kenneth. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stevens, Kenneth and Arthur House. 1955. Development of a Quantitative Description of Vowel Articulation. *Journal of the Acoustical Society of America* 27(3):484–493.

Sylak, John C. 2011. Pharyngealization in Chechen is Gutturalization. UC Berkeley Qualifying Paper. `http://linguistics.berkeley.edu/∼sylak`.

Yeou, Mohamed. 2001. Pharyngealization in Arabic: Modelling, Acoustic Analysis, Airflow and Perception. *Revue de La Faculté des Lettres El Jadida* 6:51–70.

Yeou, Mohamed and Shinji Maeda. 1995. Pharyngeal and Uvular Consonants Are Approximants: An Acoustic Modelling Study. In *Proceedings of the XIII International Congress of Phonetic Sciences*, volume 2, 409–414, KTH and Stockholm University.

John Sylak
University of California, Berkeley
Department of Linguistics
1203 Dwinelle Hall
Berkeley, CA 94720-2650

sylak@berkeley.edu

Khevsur and Tush and the Status of Unusual Phenomena in Corpora

THOMAS R. WIER
University of Chicago

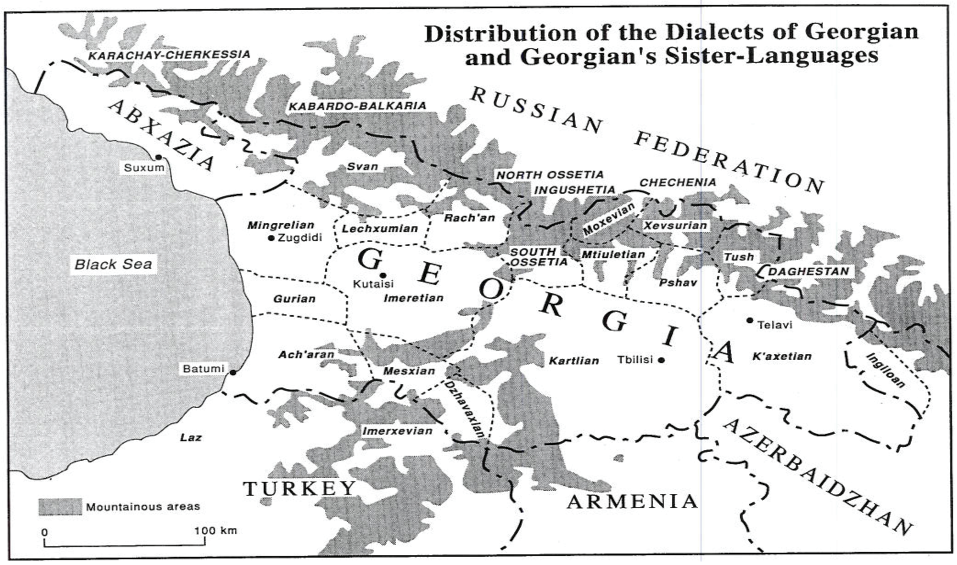
Introduction

Recent years have seen an increasing realization of the threat posed by language loss where, according to some estimates, upwards of ninety percent of all languages may go extinct within the next century (Nettle & Romaine 2002). What is less often realized, much less discussed, is the extent to which linguistic diversity that falls within the threshold of mutual intelligibility is also diminishing. This is especially true of regions where one particular language variety is both widely spoken and holds especially high prestige across many different social classes and communities. In this paper, we will examine two such 'dialects' of Georgian: Khevsur and Tush, and investigate what corpora-based dialectology can tell us about phylogenetic and typological rarities found in such language varieties.

1 Ethnolinguistic Background

Spoken high in the eastern Caucasus mountains along the border with Chechnya and Ingushetia inside the Russian Federation, for many centuries, Khevsur and Tush have been highly divergent dialects of Georgian, perhaps separate languages, bearing a relationship to literary Georgian not unlike that of Swiss German and *Hochdeutsch* (see map, from Hewitt 1995:vi).

Figure 1. Distribution of Kartvelian dialects and languages.



Among other reasons, for the better part of a thousand years they have had more intimate regular contact with Nakh-Daghestanian languages, and have borrowed numerous words from them: *balği* 'child' (cf. Lezgian *bal'a* 'child'), *bage* 'lip' (cf. Ingush and Chechen *baga*, Batsbi *bak* all 'mouth'), *ali* 'flame' (Ingush *ala*, Chechen *älu*, both 'flame'), *riq'e* 'stone' (cf. Botlikh *req'a* 'hill'). Independent of this, they have also developed a number of features not directly attributable to language contact, such as distinct lexical items (kood 'completely' instead of Standard sruliad, mtliad), differential suppletion patterns (*mi-ol* 'I'm going' vs. Standard *m-v-di-var* 'id.'), semantic shifts (Khevsur *xoq'ana* 'people' < *kveq'ana* 'land'; cf. standard *xalxi* 'people'), a different number system (*xut-oci* 'five-twenty' = 'hundred', vs. standard *asi*).

On the other hand, both dialects preserve archaisms that have been lost in some or all of the other contemporary Georgian dialects. For example, both dialects preserve a contrast between an aspirated and a glottalized uvular stop /qh/ vs. /q'/ and between the glide /y/ and /i/; the former in each case has merged uniformly with the latter in the Standard. Both dialects also preserve the permansive tense-aspect morphology and syntax from Old Georgian, a kind of gnomic aorist representing events that are always true; this has been lost in other dialects, including the standard. All of these difference add up to a significant barrier in communication for most speakers of the Standard language.

2 Previous Corpora and the Current Study

Previous work on Khevsur and Tush[1] focused primarily on lexicography, phonological processes particular to these dialects, and idiosyncrasies of paradigm formation. While like all basic documentation work this is unquestionably valuable, and much of this work has been of very high quality, a number of problems recur throughout these texts which stand in need of improvement. Firstly, many of these dialect texts were collected more than a century ago without the aid of modern recording devices and methods of elicitation. Most such texts were transcribed by hand on site, while fragile wax-cylinders and records (to the extent they ever existed) suffered the vicissitudes of neglect and outright destruction during Georgia's complex history in the twentieth century. Furthermore, metadata about the consultants' age, sex, location and relationship to the wider community were rarely or only incompletely recorded, thus making our task of interpretation all the harder. The corpora were, without exception, published in Georgian script with all commentary and linguistic analysis in literary Georgian, with the result that these dialects (or languages) were essentially inaccessible to all non-Kartvelologists. Thus scholars working on unrelated but geographically close Nakh-Daghestanian, Abkhaz-Adyghean, Turkic, Indo-European and other languages were incapable of comparing how this small area interacted within the larger ethnolinguistic context. Above and beyond these problems, however, because the dialects themselves have in all likelihood been greatly restructured in the direction of the standard language, or replaced by some sort of *Umgangsprache*, it is difficult to know whether recordings and elicitations made today are capturing the 'same' language form as that recorded a century ago. Given that any dialect of Georgian, whatever its form, also generally lies at one extreme of complexity in terms of morphosyntax among the world's languages, even specialists can have a hard time penetrating the labyrinthine relationships between paradigms, argument structure, and clausal architecture.

1 Shanidze (1984), Dolidze (1975), Chincharauli (1960)

Figure 2. Example Khevsur text:

Or-n	mama-švil-n	q'opil-an.	švil-is-ad[1]
Two-NOM.PL	father-child-NOM.PL	be.PERF-3PL	child-GEN-ADV

mama-s	col-mo-u-q'van-a=v[2].	i	kal-s	ksl-is
father-DAT	wife-PVB-PRIV-have.ANIM=QUOT	this	wife-DAT	warp-GEN

ks-ov-a	da-u-c'q'-a=v,	ksel
weave-TH-MAS.NOM	PVB-PRV-begin-AOR3SG=QUOT	warp

da-u-ks-a=v
PVB-PRV-weave-AOR3SG=QUOT

ADDITIONAL REMARKS:
1. Double case. Generate list of other instances of: -GEN-ADV
2. Noun incorporation. Generate list of other instances of: N-V (type 1).

FREE TRANSLATION:
There were two children (with a father). It's said their father had a wife for his children. This woman/wife supposedly began to weave a warp (on the loom), it's said she wove the warp.

(This text taken from Shanidze 1984)

The current study seeks to correct some of these problems by making full use of modern technology and approaches through an online digital dialect corpus. This gateway, modelled in part on the Perseus Project at Tufts University and Jost Gippert's TITUS-Projekt in Frankfurt, when completed will gather and present glossed and translated dialect texts in Georgian and Latin script in which each word is hypertexted to a dialect dictionary allowing scholars to see the cloud of meanings a given lexical entry may have. Beyond this textual level however the corpus envisions both intratextual and intertextual metatextual annotations of how a given text relates to the language and other texts in the corpus. Thus, intratextually, constructions which vary from standard Tbilisi Georgian, or from typologically expected norms, will be flagged to allow scholars unused to the norms of Kartvelian to focus on and potentially explain such differences.

Intertextually, constructions and forms in a text which differ from other texts in the corpus, either by different speakers, recorded in different locations or from different time periods, will be marked as such. This dual approach will allow scholars to see how all a given form behaves across a variety of constructural contexts. Furthermore, the digital recordings (both audio and video, where available) from which these texts were made will be made available along with each text, so that users can actually isolate the constructions in context. The goal is to give corpus users the fullest possible understanding of language use from a variety of different perspectives.

3 Typological Rara in the Corpus

3.1 Violations of Superiority Effects in Wh-Constructions

Such corpora tend to be 'messy' in the sense that they lack the idealization that accretes around studies based entirely on elicitation. Not only is this true of the current corpus of Khevsur and Tush, it reveals violations of typological norms not generally found in the already outré standard Georgian morphosyntactic system. So, for example, standard Georgian abides by the linguistic tendency that in constructions involving multiple wh-words an animacy restriction constrains otherwise rather free wordorder. In all varieties, such wh-words must surface preverbally (1a), and when both an animate and an inanimate wh-word are present, the animate wh-form must precede the latter (1b-c; Harris 1981:xx):

(1) a. ra-s a-k'et-eb-s (*ras) (St. Geo.)
what-DAT PRV-do-TH-3SG
'What is he doing?'
b. vin ra-s a-k'et-eb-s
who.NOM what-DAT PRV-do-TH-3SG
'Who is doing what?'
c. *ra-s vin a-k'et-eb-s
what-DAT who.NOM PRV-do-TH-3SG
'Who is doing what?'

Specialists who work on question constructions must often rely on elicitation because of the extreme rarity of multiple wh-constructions in corpora of natural languages. In the current corpus, however, not only are there numerous wh-constructions (or at least, more than expected from a corpus of considerably less than a million words), these multiple wh-constructions exhibit contrary tendencies in comparison with the standard dialect. As you can see in (2-3), in Khevsur, the wh-words still obligatorily surface before the verb complex (including negators). This is expected if all focal items surface immediately preverbally, as in Standard Georgian. However, other when you get two or more wh-words together, violations of superiority occur if one or more of the wh-words does not have a question interpretation, but rather is a homophonous indefinite pronoun (4-5).

(2) ra-a(d) v-k'l-av-t=av ar=c as=av čven
What-ADV 1-kill-TH-PL=QUOT not=and be.3Sg=QUOT 1PL
dana-i=v, ra=ze v-e-k'id-eb-i=ao=da
knife-NOM=QUOT what.DAT=on 1-PRV-hang-TH-PF-QUOT=and
ga-u-šv-es.
PVB-PRV-wipe-AOR3PL
'Why are we killing it, it's not our knife, what will we hang it on and what will they wipe it with?' (Kh)

(3)	peq^{h}-t	ra-s	ča-v-i-c-om-d-i=v?
	foot-DAT.PL	what-DAT	PVB-1-PRV-fall-TH-IMPF-1/2IMPF=QUOT

‘How could I fall flat on my feet?’ (Kh)

(4)	čem-tan-it	ro	c’a-xv-av,	rom	ra-s	vin
	1SG-with-INST	if	PVB-roll.up-TH	that	what-DAT	who.NOM
	vis	s-tx-ov-d-as=av				
	who.DAT	3-ask-TH-IMPF-3SG=QUOT				

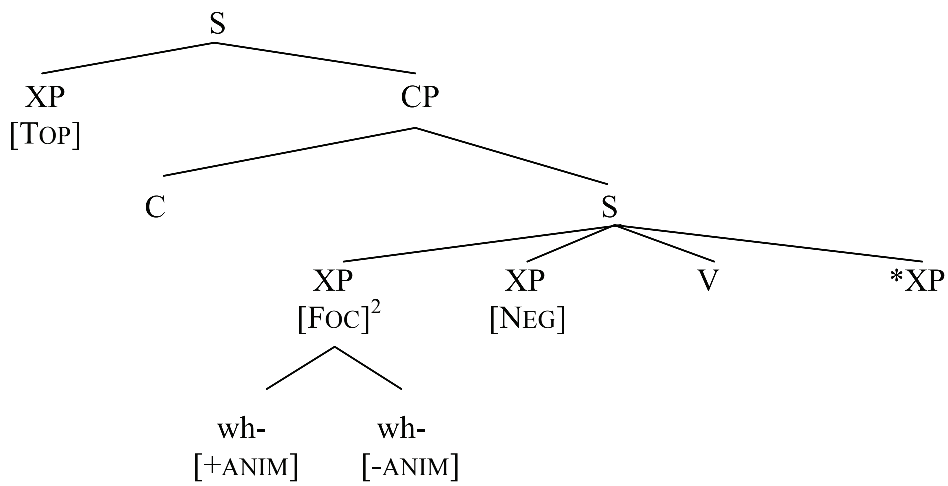
‘If you will roll it up for me so that whatever [lit. ‘what’] anyone [lit. ‘who’] asks of anyone [lit. ‘of whom’]…’ (Kh)

(5)	šen	dana	ra=ši	vis
	2SgPoss	knife.NOM	what=in	who.DAT
	š-č’ir-d-eb-od-a=v?			
	3-need-INGR-TH-COND-COND3SG=QUOT			

‘As for your knife, why would anyone [lit. ‘who’] need?’ (Tu)

This is interesting in that the morphological signaling that is usually required to obviate the underlying question interpretation to produce an indefinite in standard Georgian and western languages is not present here. Thus the templatic constraint that is present in standard Georgian (6; see Wier, forthcoming), which like these dialects also has a nonconfigurational clause structure, is weakened, in that variation in wh-word ordering occurs which does not in the standard (7).

(6) Standard Georgian clause structure



(7) Khevsur and Tush focal structure:

XP
[FOC]

That is, in comparison with the standard, the dialectal forms simply lack any templatic specification for ordering generalizations of wh-words as long as the wh-forms are all grouped together.

3.2 Noun Incorporation

In standard Georgian, noun incorporation is at best a marginal morphological process, allowable only about to the same extent that it is in English with N-N compounds:

(8) a. tav-mo-q'var-e (St. Geo.)
head-[AGENT-love-AGENT]
'proud' (lit. 'head-lover')
b. c'qal-c'a-ğ-eb-ul-i
water-PVB-take-TH-PART-NOM
'a drowning person' (lit. taken by water) (Shanidze 1953:162)

---

[2] For purposes of this article, I will remain ambiguous as to whether this is a grammaticalized syntactic feature which both wh-words with question interpretation and indefinite pronouns bear and which thus triggers particular orderings in the syntax, or whether [FOC] is a semantic or discourse functional feature. I believe it is probably the former, but this is really an empirical question testable by discourse analysis and beyond the scope of this article.

These constitute the first of the by now familiar four-way typology for noun incorporation posited by Mithun (1984, 1986):

(9) a. Type 1, Lexical compounding: heads reduce their valence by one, as in N + N > N; V + V > V; N + V > V; etc.
b. Type 2, Manipulation of case: syntactic heads not only reduce their valence by one, another argument moves in to take its place.
c. Type 3, Manipulation of discourse structure: heads (usually verbs) incorporate their dependents (usually nouns) to background incidental information.
d. Type 4, Classificatory NI: dependents incorporate into heads to act as classifiers of a more general free dependent.

In contrast to the standard language, in Khevsur, there are a variety of examples of NI, including at least one textual attestation of Mithun's Type 4 NI (aka 'syntactic' NI). In the form in (10), for example, the root *elam-* 'squint' has been incorporated into the verbal root *q'opil* 'be'[3]. A number of different criteria suggest that this form has truly been incorporated. First, this particular example comes from a story that was elicited by Č'inč'arauli who was a native speaker of Khevsur dialect. The fact that a native speaker intuitively sees them as a prosodic unit suggests (though does not prove) that they are also a morphological unit. More direct evidence of this is that the accent shifts to mark the noun as part of the verbal prosodic phrase: thus the nominative suffix *–i* receives accent in *elam-í-* instead of the initial syllable as in the free word: *élam-i*.

(10) Type 1: noun compounding

i	kal	elam-í-q'opil.
that	woman	squint-NOM-be.PERF

'The woman had a squinty-eye.'

Another argument that these arguments are truly incorporated into the verb is that the focal elements, which in standard Georgian must usually immediately precede the verb complex (excluding negators), here precede the incorporated noun:

(11)

a-dg-a	da	c'a-ma-vid,	ra	met'
PRV-stand-AOR3SG	and	PVB-vent-GO.AOR3	what	more
gza-í=a-kv…				
way-NOM=PRV-have				

'He got up and went off, however much more he had to go…' (Kh)

Perhaps most interestingly, there are even examples of Mithun's Type 4 NI in both Khevsur and Tush dialects. In the Khevsur example in (12), the incorporated

[3] Cf. standard literary Georgian: 'is kali elami iq'o'.

nominal *mrudi* ‘crooked’ modifies the more general nominal *k’utxi* ‘corner’, classifying it.

Type 4: ‘syntactic’ NI

(12)	ert	k’utx-i=v	mrud-í=akv=av
	one	corner-NOM=QUOT	crooked-NOM=have.INAN=QUOT
	‘[The house is good but it] has one crooked corner’		

In (13), on the other hand, we see evidence that the incorporated element need not be less general than the nonincorporated element. The incorporated nominal, *kali* ‘woman’ is modified by an external adjectival nominal *ukmro* ‘husbandless, unmarried’.

(13)	Im	saxl=ši	ert	u-kmr-o	kal-í-q’opil
	that.OBL	house=in	one	PRIV-husband-PRIV	woman-NOM-be.PERF
	axalgazda	kal-i,	i	kal-s	u-k’itx-a=v
	young.NOM	woman-NOM	that	woman-DAT	PRV-ask.AOR-AOR3SG=QUOT
	‘In that house there was a certain unmarried woman, a young woman, and [the old man] asked her…’				

What is more interesting, this is actually an example of subject incorporation – rather a rare phenomenon crosslinguistically (Baker 1988, Spencer 1995). Although such constructions are by no means unattested crosslinguistically, within Kartvelian they are asymptotically rare, so their relative productivity in these more conservative mountain dialects/languages reinforces the importance of the study of less prestigious varieties of ‘exotic’ languages along with standard or more widespread varieties.

3.3 Suffixaufnahme, or Double-Case?

Another unusual property of these dialects that distinguishes them from standard Tbilisi Georgian is the use of double case constructions which, however, do not necessarily take part in any system of agreement. One basic kind of construction involves the use of a genitive followed by a dative, which may be in agreement with another dative marked head noun in the same clause:

(14)	ša-x-q’var-d-a	im	col-s	im-isa-s
	PVB-3-love-IMPF-3SG	that.DAT	wife-DAT	that-GEN-DAT
	‘He fell in love with [the other man’s] wife.’			(Kh)

Here, the genitival possessor *imisas* ‘his’ involves both a genitival suffix and a dative suffix to indicate the grammatical function of the possessum, here a dative-marked direct object *cols* ‘wife’. This represents a conservative retention of an

Old Georgian *Suffixaufnahme*, whereby all genitives had to agree in case and number with the possessum, as in (15) where the *samebisa-* 'of the trinity' takes an additional instrumental case *–yta* to indicate it modifies *šec'evn-ita* 'help-INST'. This construction still exists marginally in modern standard Georgian, but only in elliptical possessive constructions where the head becomes elided, as in (16).

(15)	šec'evn-ita	c'[mid]isa	sam-eb-isa-yta	(493 A.D.)
	help-INST	holy-GEN.SG	three-COLL-GEN.SG-INST.SG	

'With the help of the Holy Trinity…' (Silogava 1994)

(16)	vis	saxl-s	e-Z-eb?	čem-i	amxanag-isa-s
	whose.DAT	house-DAT	PRV-seek-TH	my-GEN	comrade-GEN-DAT

'Whose house are you looking for? My friend's.' (Boeder 2003:46)

Although rare in the standard, such constructions are quite common in both Khevsur and Tush; among the current texts in the corpus, at least 22 *–isa-s* constructions occur in the Khevsur corpus (~10k words) and 6 times in the Tush corpus (~40k words). Less expected however are double case constructions which do not take part in any kind of agreement with a nominal head. There are a variety of different kinds of double case, including genitive+dative *-isa-s* (where in its non-agreeing manifestation it usually functions as an adjunct), genitive+nominative *–is-i*, genitive+instrumental *–is-it*, genitive+adverbial *–is-ad*, and, exceptionally, double instrumental *–it-it*. Although almost all of these make use of a genitival stem plus some oblique case, it is unclear that the genitive contributes any meaning to the form; rather it seems simply to serve as the building block onto which further case forms (themselves rarely bearing a consistent meaning) can attach. For example, a double genitive+adverbial frequently reflects a thematic recipient of *verba dicendi* as in (17) and (18) or *verba sentiendi* as in (19) and (20), but sometimes merely the experiencer (19), and sometimes the thing being experienced (20). Finally, sometimes the double-case form marks the recipient, as in (21).

(17)	"peq^{h}-t	ra-s	ča-v-i-c-om-d-i=v?"
	foot-DAT.PL	what-DAT	PVB-1-PRV-fall-TH-IMPF-1/2IMPF=QUOT
	u-tkv-am-is	memcxvar-is-ad	
	PRV-say-TH-3SG	shepherd-GEN-ADV	

'"How could I fall flat on my feet?" he says to the shepherd.' (Kh)

(18)	diac-is-ad	u-ʒex-eb-a=v			"k'arg oq^{h}šam, k'arg
	peasant.woman-GEN-ADV	PRV-call-TH-3SG=QUOT			good strap good
	mo-gv-i-mzad-e=v		me	da	čem st'umar-sa=v!"
	PVB-1PL-PRV-prepare-AOR1/2=QUOT		1SG	and	my guest-DAT=QUOT

'He calls the peasant woman: 'Prepare a good strap – a good one! – for me and my guest!" (Kh)

(19) im cxvar-s k'ide su da-ckmat'ur-eb-ul-sa
that.DAT sheep-DAT again just PVB-?-TH-PART-DAT
im-is-ad e-xed-v
that-GEN-ADV PRV-see.IMPF-TH
'The sheep again made [unclear] had just caught sight of him.' (Tu)

(20) u-t'ir-is=ad memr gada-k'id-eb-ul-iq'v kmr-is-ad
PRV-cry-3SG=and then PVB-irritate-TH-PART-3SGPF husband-GEN-ADV
'She cried and bugged her husband…" (Tu)

(21) memr is ak'avan gē-y-k'et-a=d dē-y-c'v-in
then that.NOM cradle PVB-PRV-do-AOR3SG=and PVB-PRV-burn-CAUS
is balğ-i. Še-mē-y-č'ir še-mē-y-č'ir
that.NOM child-NOM PVB-VENT-PRV-cut PVB-VENT-PRV-cut
im ak'avan-s Še-mē-y-č'ir še-mē-y-č'ir
that.OBL cradle-DAT PVB-VENT-PRV-cut PVB-VENT-PRV-cut
im ak'avan-s im balğ-is-ad,
that.OBL cradle-DAT that.OBL child-GEN-ADV
še-mē-y-č'ir še-mē-y-č'ir-a=d mē-y-k'l
PVB-VENT-PRV-cut PVB-VENT-PRV-cut-Aor3Sg=and PVB-PRV-die
'Then he made the cradle and burned the little child. He dashed and smashed on the cradle, he dashed and smashed on the cradle for the child, and he dashed and smashed and he [the child] died.' (Tu)

It's also worth pointing that while such double-case forms may encode obligatory arguments, as above, they can also encode optional adjuncts, as in (22) and (23) below.

(22) im-is udumliv mamamtl-is ksl-is-ad
this-GEN stealthily father.in.law-GEN weave-GEN-ADV
gara ga-mo-u-c'vd-av-a=d da-u-mal-av
heddle.stick PVB-VENT-PRV-reach-TH-EP=and PVB-PRV-hide-th
'She stealthily reaches for her father-in-law's heddle-stick for weaving and hides it from him.' (Kh)

(23)	im	beber[4]	diāyc-s	k'ide	q^{h}el
	that.OBL	old.man	old.woman-DAT	again	hand
	c'a-mē-y-vl-ū		im	napot'-is-ada=d	
	PVB-VENT-PRV-seize-AOR3SG		that	roach-GEN-ADV=and	
	c'ē-y-ğ-o=d		k'idoban=ši	čē-y-d-v.	
	PVB-PRV-take-AOR3SG=and		bin=in	PVB-PRV-put-AOR3SG	

'The old man took the woman by the hand and brought it to the roach and put it [the roach] in the bin.' (Tu)

The genitive+adverbial double-case construction predominate throughout the texts. However, as noted above, most of the other logically possible combinations of genitive plus another case are possible. In (24), we see a postposed genitival modifier *k'ac-is-i* 'man-GEN-NOM' agreeing in case with the head noun *gon-i* 'thought-NOM'[5].

***-is-i* [GEN-NOM]**

(24)	gon-i	k'ac-is-i	m-kon-d-a,	magre
	thought	man-GEN-NOM	1SG-have.INAN.IMPF-IMPF-3SG	but
	ʒağl-i	v-i-q'av-i		
	dog-NOM	1-PRV-be.IMPF-1/2		

'I had the thought of a man, but I was a dog' (Kh)

In (25), we see a genitival form *dğ-is-ita* 'day-GEN-INST' which functions as an adjunct of time:

***-is-it* [GEN-INST]**

(25)	c'a-vid-o=d	is	k'ac-i
	PVB-go.AOR-OPT3SG=and	this.NOM	man-NOM
	da-i-c'q'-eb-d-a	ʒal-s.	dğ-is-ita=c
	PVB-PRV-begin-TH-IMPF-IMPF3SG	force-DAT	day-GEN-INST=too
	ša-i-ʒl-eb-od		
	PVB-PRV-can-TH-COND		

'The man went, [and] really got a start. By day he was able.' (Kh)

Finally, we also find examples of multiple exponence of the same case, as in (26), where one instrumental case suffix is followed by another identical case suffix

4 Interestingly, this argument lacks the expected narrative case suffix *–m* (standard Georgian *–ma*). It does however use the oblique form of the article used for the narrative case, rather than the nominative case of the article *is*. This is another example of a shift from head-marking to dependent-marking in these dialects.

5 In standard such postposed modifiers are, rarely, possible in high literary style, but they would be constructed differently in the nominative: *goni k'ac-isa*, where the suffix *–isa* is the extended form of the genitive from Old Georgian used in relics like this.

apparently with no difference in meaning from the standard form *ik-it* [there-INST] 'over there':

***-it-it* [INST-INST]**

(26) im dro-s i mamamtl-is xalx
that.OBL time-DAT that father.in.law-GEN people
mo-svl-i=v cxra-mta-s <u>ik-it-it</u>:
PVB-go.MAS-NOM=QUOT nine-mountain-DAT there-INST-INST
'At that time, her father-in-law's people were going <u>up there</u> to Nine-Mountain.' (Kh)

What is interesting about most of these double-case constructions is that they do not fit into standard discussions of the phenomenon of *Suffixaufnahme*, in which dependents must agree with heads in case and perhaps other features for the simple reason that in most of the above cases there is never any evidence of a head noun with which the nominal in question could stand in agreement. Even when the doubly-marked noun serves discourse-functionally as a modifier, as in (21), it does always do so syntactically, as when *balğ-<u>is-ad</u>* [child-GEN-ADV] 'for the child' stands as a sentential adjunct and the extra case does not agree with that which it discourse-functionally modifies, *ak'avan-s* [cradle-DAT].

An immediate question thus arises: if these doubly marked constructions do not uniformly arise from some process of agreement, as they would have in Old Georgian, where do they come from? One intriguing possibility is that they result from language contact with Nakh languages to the immediate north of the Khevsur and Tush dialect regions in Georgia. As noted above, these dialects have for many centuries been in contact with Chechen, Ingush, Bats, and other North Caucasian languages, which have the unusual property that oblique cases do not attach directly to the nominal root, but rather to a stem formant exclusively used for obliques. These oblique stem formants vary both in form and distribution across Nakh-Daghestanian languages, but typically behave much like the Archi and Batsbi forms in (27) and (28) below:

(27) Archi (Diana Forker p.c.)
a. gel
cup.ABS.SG
b. gel-li-s
cup-OBL-DAT

(28) Tsova-Tush (Holisky and Gagua 1994)

	'bear'	'broom'
NOM	ča	koǯ-o
GEN	ča-i-^{n}	koǯ-ni-^{n}
DAT	ča-i-n	koǯ-ni-n
ERG/INST	ča-i-v	koǯ-ni-v
CON	ča-i-x	koǯ-ni-x
ALL	ča-i-go	koǯ-ni-go
ADV	ča-i-ğ	koǯ-ni-ğ
COM	ča-i-ci^{n}	koǯ-ni-ci^{n}

Nominals having distinct oblique stems are not entirely rare; they occur e.g. in well-known Indo-European languages like Latin and Greek. The interest in these Nakh-Daghestanian forms lies in the fact that the oblique formant is separable but semantically vacuous from the point of view of case features. That is, although they sometimes bear other features like the singular/plural contrast in the Batsbi forms, the oblique formant serves only to provide a licit stem for the actual case suffix. It is this fact that brings us back to a discussion of Khevsur and Tush double-case, since the genitival suffix in most instances is semantically vacuous in exactly the same way, with the difference that that most of the oblique cases (i.e., those that are not nominative in the Kartvelian context) may also surface attached directly to the nominal root without any oblique stem intermediary. Thus, while the system is not identical to any Nakh-Daghestanian one, it is as if the Khevsur and Tush speakers, many of whom were also presumably bilingual in one or more other Nakh-Daghestanian language, were borrowing a constructional device from those languages but using their own indigenous resources, *Suffixaufnahme*, which had originally served a quite different function, to do so.

4 Conclusion

This survey of properties of Khevsur and Tush dialects of Georgian has shown the value of data-focused corpus studies for studies of linguistics and typology, since they have a tendency to confound traditional notions of how grammars are supposed to work. Khevsur and Tush show that even when a standard form of the language abides by supposed notions of superiority in wh-constructions, some dialects can and do violate these norms. They also show that noun incorporation may indeed vary, and even unusual forms of noun incorporation such as subject NI or syntactic NI may occur in one variety while in another variety NI is almost completely absent. Finally, most interestingly, even in languages noted for obscure construction types, such as Suffixaufnahme or double-case, nonstandard varieties of language may contain typologically unusual variants of those same construction types, rarities within rarities.

References

Baker, Mark. 1988. *Incorporation: a theory of grammatical-function changing*. Chicago: University of Chicago Press.

Ch'inch'arauli, Aleksandre. 1960. *Xevsurulis taviseburebani: t'ekst'ebita da indeksit*. [Peculiarities of Khevsur: with texts and an index]. Tbilisi: Press of the Georgian SSR Science Academy.

Dolidze, Givi. 1975. *Xevsuruli t'ekst'ebi: leksik'onebiturt* [Khevsur texts: along with lexica]. Tbilisi: TSU Press.

Harris, Alice C. 1981. *Georgian Syntax*. Cambridge: CUP.

Hewitt, S. George. 1995. *Georgian: a structural reference grammar*. Philadelphia: John Benjamins.

Holisky, Dee, and Rusudan Gagua. 'Tsova-Tush (Batsbi)' in *The Indigenous Languages of the Caucasus, 4: North East Caucasian, Part 2*, ed. by Rieks Smeets, 147-212. Delmar, N.Y.: Caravan Press.

Mithun, Marianne. 1984. 'The Evolution of Noun-Incorporation.' *Language* (60:4): 847-894.

Mithun, Marianne. 1986. 'On the Nature of Noun-Incorporation.' *Language* (62:1): 32-37.

Nettle, Daniel and Suzanne Romaine. 2000. *Vanishing-Voices: the Extinction of the World's Languages*. Oxford: OUP.

Shanidze, Ak'ak'i. 1953. *Kartuli gramat'ik'is sapuʒvlebi* [Foundations of Georgian Grammar]. Tbilisi: TSU Press.

Shanidze, Ak'ak'i. 1984. *Txzulebani tormet' t'omad, Vol. 1: kartuli k'iloebi mtaši*. [Compositions in twelve volumes: Georgian dialects in the mountains]. Tbilisi: TSU Press.

Dr. Thomas R. Wier
Dept. of Linguistics
1010 E. 59th St.
Chicago, IL 60615

trwier@uchicago.edu