Ejective harmony in Lezgian

Avery Ozburn
University of British Columbia

Alexei Kochetov
University of Toronto

Supplementary materials

Appendix A: Affixes: stress and potential harmony triggers and targets

1 Lists of affixes by stress type

As discussed in §3.2, affixes in Lezgian can be stress-attracting (34a) or stress-neutral (34b). Affixes of the first type cause a stress shift from the stem; affixes of the second type do not affect stress in the stem, because they are always unstressed. Some stress-attracting affixes attach to monosyllabic roots, forming a regular binary foot (e.g. [tar-'ar] 'trees'); other stress-attracting affixes can attach to roots/stems of any length, thus causing exceptional stress patterns: on the first (['bej-wafa] 'treacherous'), third ([xabar-'dar] 'informed') or fourth syllable ([merhemet^h-'suz] 'merciless').

(34) a. *Stress-attracting affixes* (Haspelmath 1993: 67–68)

	i.	*Affixes forming a default foot* (*with monosyllabic roots*)

-Ar, -lAr	(plural)
-rA, -A, -A-di, -Uni, -di [ʦi]	(oblique stem)
-A, -U	(verbal theme)
-U	(adjective-deriving)
-Aʦ^h	(noun-deriving)
-ji	(aorist participle)

ii. *Affixes forming an exceptional foot* (*with roots of any length*)

-lu (loan)	(adjective-deriving)
-suz (loan)	(adjective-deriving)
-dar (loan)	(adjective-deriving)
-k^{h}ar (loan)	(adjective-deriving)
bej-	(adjective-deriving)

b. *Stress-neutral affixes* (Haspelmath 1993: 69)

-ar	(plural)
-di, -i	(oblique stem)
-aj	(elative)
-iz	(infinitive)
-un	(masdar)
-na	(aorist)
-z(a)wa	(imperfective)
-n(a)wa	(perfect)
-da	(future)
-raj	(optative)
-mir	(prohibitative)
-a	(imperative)
-aj	(aorist participle)
-la	(temporal converb)
-daldi	(posterior converb)
-n(a)maz(di)	(immediate-anterior converb)
-lda	(hearsay evidential)
-ni	(interrogative)
-t'a	(conditional)
-ni	(additive suffix)
-wal	(abstract noun)
-ra	(multiplicative)
-dakʰaz, -diz	(deadjectival adverb)
-tʃʰi	(agent suffix)
-xana	(house suffix)
-ar	(causative verb-deriving)
tA-, tU-	(negative)
Aq'-, Aq'al-, Aʦ-, Ak-, etc.	(preverb prefixes (see below))

It should be noted that some plural and oblique stem suffixes have stress-attracting allomorphs (e.g. /-Ar/, /-di/ [ʦi]) and stress-neutral allomorphs (e.g. /-ar/, /-di/). While these are partly in complementary distribution with respect to the type of roots – monosyllabic *vs.* polysyllabic – stress-neutral allomorphs can attach to a few monosyllabic roots. These are either loanwords (e.g. ['tʃʰan-di] 'soul' (< Persian)) or roots that were historically disyllabic (e.g. ['tʃʰkal-di] 'peel').

2 Affixes with ejectives and plain stops

The following is a list of Lezgian affixes with ejectives and plain voiceless stops (other than the harmonising suffix /-di/), based on Haspelmath (1993: 557). Note that there are only five affixes (three prefixes and two suffixes) with ejectives, and eight affixes, all prefixes, with plain voiceless stops. The lack of suffixes with plain voiceless stops is probably related to the restriction on these segments in posttonic contexts. Except for /-t'a/, all of these affixes are limited to a closed set of forms or a few words. Furthermore, the preverbs listed here, which comprise all of the prefixes except the negative /tA-/, are described by Haspelmath (1993: 167) as morphologically opaque and as not being analysed as decomposable by native speakers. In this case, there are no prefixes with underlying ejectives, and only one with an underlying plain voiceless stop.

	prefixes		suffixes	
a.	Aq'-, Aq'al-, kʷ'-	preverb denoting localisation (a few words)	-k'a	adjectival adverb (closed set)
			-t'a	conditional
b.	tA-	negative (closed set)	*n/a*	
	Aʦ-, Ak-, Aq-, kAk-; kʷ-, qAq-, χk	preverb denoting localisation (a few words)		

Table VI
A list of Lezgian affixes with (a) ejective and (b) plain voiceless
stops (based on Haspelmath 1993: 557; see pp. 114–115, 132,
133, 170–174 for descriptions of specific affixes).

Our subsequent search through Haspelmath's (1993) grammar revealed no examples of the prefixes containing ejectives, namely /Aq'-/, /Aq'al-/ or /kʷ'-/, followed by plain voiceless stops, nor any of the ejective suffix /-k'a/ preceded by plain voiceless stops (but some examples of these affixes with other stops, e.g. [aˈq'al-tʰun] 'go up, appear', [aˈq'al-t'un] 'finish', [aq'-ˈadarun] 'put in', [qʷ'-ˈadarun] 'shake off', [tʃ'uˈru-k'a] 'wrongly', [ʦ'iˈji-k'a] 'newly'). However, there were multiple examples of the productive ejective suffix /-t'a/ attaching to stems containing plain voiceless stops, some of which are shown in (35a). For the plain voiceless prefixes, both the preverbs and the negative /tA-/ do attach to stems with ejectives. The negative prefix, as discussed in the paper (see (13) in §4.3), shows optional laryngeal dissimilation. Examples with preverb prefixes are shown in (b) (based on Haspelmath 1993).

(35) *Sample forms with ejective and plain voiceless stops*

    a. ˈtir-t'a   'to be (PAST, COND)'

        aˈku-r-t'a   'to see (AORIST PARTICIPLE, COND)'

        aˈqatʰ-aj-t'a   'to fall (AORIST PARTICIPLE, COND)'

        t-aˈwu-r-t'a-ni   'even to do (NEG, AORIST PARTICIPLE, COND)'

        aˈta-na-t'a   'to come (AORIST, COND)'

        ta-x-ˈaj-t'a   'if it is not, otherwise'

        ˈxta-j-t'a-ni   'even to return (AORIST PARTICIPLE, COND)'

        te-ˈfe-j-t'a-ni   'even to go (NEG, AORIST PARTICIPLE, COND)'

    b. ats-ˈuq'un   'to sit down'

        eq-ˈetʃ'-un   'go out'

        ek-ˈetʃ'-un   'to join'

        qeq-ˈetʃ'-un   'to go away'

        χk-ˈetʃ'un   'to go out'

        ku-ˈmuq'un   'to remain under'

In (35a), as mentioned, the ejective suffixes are stress-neutral. As such, the ejectives consistently occur outside the foot. For example, the foot structure of the first example in (a) is ('tir).t'a; as such, even though the plain voiceless stop and ejective are onsets in adjacent syllables, the ejective is outside the foot, hence outside the domain of harmony, and therefore cannot trigger harmony on the plain voiceless stop in the root. In other words, given the foot-sensitive domain of harmony, we do not expect the process to apply here.

    Turning to the prefixes in (35b), all are derivational preverb constructions that denote localisation (cf. [aˈqal-tʰun] 'to go up, appear', [aˈqal-t'un] 'to finish'). As mentioned earlier, Haspelmath (1993: 167; see also 170–174) describes preverbs as morphologically opaque forms that are no longer decomposable by native speakers. Given this, preverbs are not particularly useful for testing heteromorphemic patterns and, instead, will be included in our lexicon analysis in Appendix B. Nevertheless, it is worth noting that, as shown in (b), these words do occasionally contain combinations of plain voiceless stops followed by ejectives. Importantly, in all of these cases, the potential trigger of harmony, namely the ejective, occurs beyond the stress and thus outside the foot. Again, given the harmony domain, we would not expect it to apply here.

Appendix B: Details of the lexical analysis methodology

The following are additional details on the creation of our lexical corpus of roots with non-local combinations of stops, based on Talibov & Gadzhiev's (1966) Lezgian–Russian dictionary.

The Cyrillic spelling of Lezgian words was converted to IPA. The software Phonological CorpusTools (PCT; Hall *et al.* 2015) was used to compile a list of words containing two or more stops/affricates. The resulting list of 4821 words was further edited manually to exclude words with local co-occurrences (adjacent stops), unassimilated loans from Russian (but not identifiable loans from other languages) and derived forms.

Criteria for loanwords. The great majority of loanwords from Russian (many of which are words of Western origin) have been introduced into the language recently, starting in the 1920s. These words are listed in the dictionary in standard Russian orthography. They were removed from the corpus. We chose not to remove the so-called 'Oriental loans' – words from Arabic, Iranian (primarily Persian, Kurdish and Tat) and Turkic (primarily Azerbaijani, Turkish and Kumyk) languages (Selimov 2001). Most of these words were introduced into the language over many centuries of either direct language contact or extensive cultural/religious influences (Talibov 1966: 601–602, Haspelmath 1993: 25–27), and have at least in part been assimilated into the phonology of Lezgian. For example, many loans from Arabic contain ejectives that are absent in the language of origin (e.g. [qʼaˈlub] 'form, pattern' < Arabic [qaalib]; Haspelmath 1993: 496). In our corpus, Oriental roots account for almost half of the data. We also kept older Russian loans which appear to be adapted (e.g. [pʼaˈpʼrus] from [papirosa] 'cigarette') and names of ethnic groups of the Caucasus (e.g. [darˈgi] 'Dargin').

Criteria for derived forms. Derived forms were defined as words containing known derivational and/or inflectional affixes (based on Haspelmath 1993 and other sources). For nouns exhibiting root alternations, we included only absolute singular forms (e.g. [tʼab] 'lie', but not the harmonic ergative singular form [tʼapʼ-ˈuni]; see (10)). Entries with preverb constructions (see the previous section), however, were kept on the list, as these are considered by Haspelmath (1993) to be morphologically opaque and semantically independent. If a given word had two variant forms with different laryngeal classes (e.g. [qaʃʼqa] ~ [qʼaʃʼqʼa] 'white spot (for horses)'), both were retained in the corpus.

With these corrections to the original PCT output, the resulting list contained 1284 words that can be assumed to accurately reflect the patterns in Lezgian roots. We manually added exceptional stress (when known), part of speech and syllable structure, and also indicated non-native roots identified using Selimov's (2001) etymological dictionary. Further, in roots with multiple non-adjacent stops, C_1–C_2–C_3..., only the co-occurrences C_1–C_2 and C_2–C_3 were included in the counts, so that the counts reflect stops that are adjacent on a stop tier. In roots that contained both adjacent and non-adjacent stops, such as C_1–C_2C_3–C_4, only the non-adjacent co-occurrences,

i.e. C_1–C_2 and C_3–C_4, were included. A total of 1418 non-local stop co-occurrences were obtained.

Among these co-occurrences, 730 (51%) were represented by native words, and 688 by (mainly Oriental) loanwords, as shown in Table VII.

native	730
foreign	688
Arabic	199
Iranian	206
Turkic	254
ethnic name	10
other	19
total	1418

Table VII
Counts of co-occurrences by lexical stratum.

Table VIII provides counts of roots by consonants. Note that ejectives (as C_1 or C_2) are the most frequent category in the native stratum. They are much less numerous in the foreign stratum; nevertheless, their presence there indicates that many loanwords are well integrated in the language (as none of the major donor languages have ejectives). In particular, voiced stops (especially as C_1) are relatively infrequent in the native vocabulary, while being highly frequent in the foreign vocabulary.

	C_1					C_2					
	D	T	T'	T^h	$T^{(h)}$	G	K	K'	K^h	$K^{(h)}$	
native	121	156	223	152	78	148	82	280	187	33	730
foreign	256	117	59	123	133	260	78	67	238	45	688
total	377	273	282	275	211	408	160	347	425	78	1418
% native	32	57	79	55	37	36	51	81	44	42	

Table VIII
Counts of co-occurrences by C1 and C2 and the native and foreign strata. $T^{(h)}$ and $K^{(h)}$ = unidentified aspirated or plain voiceless stop.

Finally, Table IX presents counts of cases by prosodic shape. Note that the most common type (43%) involves co-occurrences of two onsets within an iambic foot.

Foot/syllable	Position	Count	Example
(w<u>s</u>)	onset–onset	614	ba'qal
(s)	onset–coda	219	kap^h
(w<u>s</u>)	onset–coda	152	ʃy'q'ynt^h
(<u>ws</u>)	onset–coda	121	k'u'luk^h
(w<u>s</u>)w	onset–onset	100	ru'ts'ugul
(<u>ws</u>)	onset–coda	57	gat^h'far
(<u>ws</u>)w	onset–onset	53	gi'liq'un
(<u>ws</u>)w	onset–coda	20	be'rek^hat^h
(<u>ws</u>)w	onset–coda	16	mu'qajat^h
(<u>ws</u>)w	onset–coda	15	ba'sarat^h
(s)<u>w</u>	onset–onset	14	'k'udun
(s)<u>w</u>	onset–onset	7	'χkat'un
(w<u>s</u>)	coda–coda	7	maq'ʼsad
(ws)<u>ww</u>	onset–onset	5	sa'rubuʁda
(s)<u>w</u>	onset–onset	2	'belki
(s)<u>ww</u>	onset–onset	2	'jaqadaʃ
(<u>ws</u>)w	coda–coda	2	matʃ^h'χulat^h
(<u>ws</u>)w	onset–onset	2	t'yry'q'ym
(<u>ws</u>)w	onset–coda	2	hu'k^humat^h
(s)	onset–onset	1	k^hstaχ
w(<u>ws</u>)	onset–coda	1	istek^h'lu
(s)w<u>w</u>	onset–onset	1	'baʃyste
(s)<u>w</u>	onset–coda	1	'helbet^h
(<u>ws</u>)w	coda–onset	1	miq'ʼnat'is
(<u>ws</u>)w	coda–onset	1	ʃib'lit^har
(w<u>s</u>)w<u>w</u>	onset–onset	1	i'ʧ^halat^har
(ws)<u>ww</u>	onset–onset	1	a'bugerden

Table IX

Co-occurrences by foot structure, participating syllables and
syllable positions; w/s = weak/strong (unstressed/stressed)
syllables; underlining indicates the location of participating stops.

Appendix C: The dictionary corpus

The corpus used for this paper, which was created from the electronic version of Talibov & Gadzhiev's (1966) Lezgian–Russian dictionary, is available at https://doi.org/10.1017/S0952675718000118, in the form of a spreadsheet.

The columns included in the spreadsheet are listed below.

A. #: a unique identifying number for each word in the corpus, based on rows being sorted alphabetically.

B. Word: IPA transcripion.

   H indicates that the presence of aspiration (which in most cases is not reflected in orthography) is unknown.

C. Transcription: non-adjacent segments on a stop tier (i.e. the list of stops), separated by periods.

   As above, H indicates that the presence of aspiration is unknown.

D. C_1 class: the laryngeal class of the first consonant in the co-occurrence.

   D = a voiced stop/affricate, T' = an ejective stop/affricate, T = a plain stop/affricate, T^h = an aspirated stop/affricate and T(h) = a stop/affricate with unknown aspiration.

E. C_2 class: the laryngeal class of the second consonant in the co-occurrence.

   The notation is the same as above, except that velars appear instead of coronals.

F. Place class: heterorganic (ht) or homorganic (hm) co-occurrence.

G. Stratum: the foreign or native status of the word (based on Selimov 2001).

H. Loan origin: for words in the foreign stratum, the language which the loan originates from (based on Selimov 2001).

I. CV tier: the CV structure of the co-occurrence domain.

   c = stop/affricate, s = fricative, r = sonorant consonant, V = a stressed vowel in polysyllabic words, v = an unstressed vowel in polysyllabic words or a stressed vowel in monosyllabic words, '.' = syllable boundary.

J. Foot type: co-occurrences classified by foot structure, participating syllables and syllable positions.

   – ft1 = both C_1 and C_2 occur in the first (or only) foot; ft2 = both C_1 and C_2 occur in the second foot; ft1-2 = C_1 is in the first foot, while C_2 is in the second foot;

   – W = a weak (unstressed) syllable; S = a strong (stressed) syllable (words are represented as sequences of W/S syllables, with hyphens indicating foot boundaries, e.g. WS-W for [aˈdalatʰ]);

   – syll1 = both C_1 and C_2 occur in the first (or only) syllable; syll2 = both C_1 and C_2 occur in the second syllable; syll1-2 = C_1 is in the first syllable, while C_2 is in the second syllable, etc.;

   – ons ons = both C_1 and C_2 are onsets; ons coda: C_1 is an onset and C_2 is a coda, etc.

Based on this classification, the top five types (accounting for 85% of the corpus) are:

1. ft1 WS - syll1-2 - ons ons e.g. [baˈqal]
2. ft1 S - syll1 - ons coda e.g. [kapʰ]
3. ft1 WS - syll2 - ons coda e.g. [ʃyʼˈqʼyntʰ]
4. ft1 WS - syll1-2 - ons coda e.g. [kʼuˈlukʰ]
5. ft1-2 WS-W - syll2-3 - ons ons e.g. [ruˈʦʼugul]

K. Part of speech: the part of speech of the word.

n = noun, v = verb, adj = adjective, adv = adverb, interj = interjection, pron = pronoun, onom = onomatopoeia, '?' = unknown.

L. Dictionary entry: the first line of the original dictionary entry from the source, with the Lezgian words transcribed in IPA.

(Alt) = one of the relevant consonants is involved in alternations in oblique forms.