

LOANWORD TYPOLOGY THE CASE OF BEZHTA

Bernard Comrie and Madzhid Khalilov
Max Planck Institute for Evolutionary Anthropology
comrie@eva.mpg.de, khalilov@eva.mpg.de

Existing hypotheses on lexical borrowing

1. Some concepts are more resistant to lexical borrowing than others especially basic vocabulary
 - cf. English *hand*, German *Hand*; English *ballet*, French *ballet*
 - but there is no scientific definition of the concept “basic vocabulary”
 - the Swadesh list purports to be a universal list of the human condition
 - but there are exceptions (e.g. ‘ice’, ‘horn’)
 - such a list of lexical concepts is a hypothesis that must be tested in a systematic and cross-linguistic manner
 - especially some semantic fields
 - but the identification of such semantic fields also needs systematic investigation
 - some semantic fields behave differently in different cultural environments
 - e.g. Austronesian languages have few loanwords connected to the sea, while (High) German has borrowed much of this vocabulary from Low German
 - especially more frequent words
 - but again there are almost no empirical investigations on the basis of a wide range of languages representing various kinds of language contact situations
 - especially some parts of speech
 - cf. the hypothesis that nouns are more borrowable than verbs, or that grammatical words are more resistant to borrowing than lexical words
2. Loanword Typology Project
 - (project leaders: Martin Haspelmath, Uri Tadmor)
 - worldwide sample of languages
 - representing different language families, areas, and types of contact
 - in this case ca. 50 languages
 - list of words must be more extensive than Swadesh
 - and include words from presumed non-basic vocabulary
 - in order to test Swadesh’s hypothesis
 - in this case a list of over 1300 words
 - as used for the “Intercontinental Dictionary Series” project
 - each language is investigated by a specialist on the given language (and its history)
 - each author submits a data base, i.e. an electronic list of the lexical items, corresponding to the given 1300 concepts
 - identifying loans and their immediate sources
 - each author also submits a chapter presenting empirical and theoretical generalizations
 - the project is made possible by recent work compiling dictionaries of various languages of the world, including comparative and etymological dictionaries
3. Results of the project
 - for each lexical concept, the percentage of languages with a loanword
 - this will give a list of lexical concepts from the most to the least borrowable
 - and facilitate the testing of hypotheses concerning the relevance of semantic field, frequency, part of speech, etc.

the comparison of different language contact situations
will facilitate the testing of hypotheses concerning the relevance of the social
environment for lexical borrowing

Bezhta

4. Bezhta (Kapuchi) [kap], ca 6461 speakers (2002 Russian census; informal estimates give around 12,000) in west of Republic of Daghestan
main dialects: Bezhta Proper, Khasharkhota, Tladal
Bezhta Proper has some idiosyncratic properties:
shift of *r to ʁ
shift of *h, *ʃ to h, ʔ respectively with umlaut of neighboring vowels

5. Nakh–Daghestanian language family:

Nakh–Daghestanian (North–East Caucasian, East Caucasian)

Nakh languages

Daghestanian languages

Avar–Andi–Tsezic languages

Avar–Andic languages

Avar [ava]

Andic languages

Tsezic languages

West Tsezic languages

Khwarshi [khv] (incl. aberrant Inkhoqwari “dialect”)

Tsez (Dido) [ddo]

Hinuq [gin]

East Tsezic languages

Bezhta (Kapuchi)

Hunzib [huz]

Lak

Dargi

Lezgif languages

Khinalug

6. Major resources, mostly from recent years
 - a. dictionaries of Bezhta, Hunzib, Tsez, and Hinuq (but not Khwarshi)
 - b. comparative dictionary of North Caucasian (Nikolayev and Starostin 1994)
7. Major lacuna
no systematic work on inter–Tsezic loans
partly in absence of detailed comparative phonology
8. Inherited vocabulary
 - a. vocabulary inherited from Proto–East Tsezic (ca 1000 BP)
 - b. vocabulary inherited from Proto–Tsezic (ca 2000 BP)
 - c. vocabulary inherited from Proto–Nakh–Daghestanian (ca 8000 BP)
 - d. in principle, also vocabulary inherited from
 - i) Proto–Avar–Andi–Tsezic
 - ii) Proto–Daghestanian
9. Major non–Tsezic linguistic contacts of Bezhta:
direct contacts:
Georgian [kat] (at least from early centuries CE)
Avar (long–standing, heavy since 1860s)
Russian [rus] (direct influence mainly from 1950s; some earlier influence through Georgian and especially Avar)

indirect contacts:

Arabic [arb] (at least from C18; assumed to be through Avar, with possible transmission to Avar via Turkic, esp. Azerbaijani [qzj] and Kumyk [kum])

Persian [prs] (assumed to be through Avar, with possible transmission to Avar via Turkic, esp. Azerbaijani and Kumyk)

Turkic, esp. Azerbaijani and Kumyk (assumed to be through Avar)

Sources of loans

10. Georgian

main studies Xalilov (1993; 2004)

loans into

Bezhta/Hunzib (ca 400 items)

Tsez/Hinuq (ca 129 items, of which 88 in Tsez or Tsez and Hinuq)

Khwarshi (ca 10 items, all nouns — less borrowing? or secondary loss?)

some predate language-specific sound changes within Tsezic, so probably reflect Proto-Tsezic period loans, e.g. Geo *gemo* 'taste', Bezhta-Hunzib *gemo*, Tsez-Hinuq *gimu*; Geo *okro* 'gold', Bezhta-Hunzib *okro* (Bezhta Proper *okko*), Tsez-Hinuq *ukru* 'money, silver'

some show typical Old or Middle Georgian forms, e.g. O/MiGeo *toqi* 'hoe', MoGeo *toxi*, Bezhta-Hunzib-Hinuq *toqi*, Tsez *toqʼi*, though most are Modern Georgian

Bezhta-Hunzib (but not Tsez-Hinuq) have some Georgian loan verbs as parts of light verb constructions (LVCs), e.g. Bezhta *dayup'a -owal* 'spoil' < Georgian *da-yup'v-a* (PFV-spoil-INFIN), also in some dialects the conjunction *magram* 'but'

loans mainly concrete nouns, reflecting Georgian cultural influence, e.g. Bezhta *leybi* 'mattress' (< *leibi*), *topi* 'rifle' (< *topi*)

11. Avar (esp. adjacent Antsukh (*ánsuq*) dialect)

traditional lingua franca of western Daghestan

at least fm 1860s, especially fm 1920s

also conduit for loans fm other languages

Avar origin vocabulary

nouns from wide range of semantic domains, e.g. *beyten* 'wedding', *waccal* 'cousin'

includes many adjectives, e.g. *bercinab* 'beautiful' (in Avar *-b* is a class III [neuter] suffix; in Bezhta, the word is invariable)

verbs are borrowed only as parts of LVCs, in the infinitive form in *-(i)zi*, e.g. *bak'arzi -owal* 'gather (tr.)', *bak'arzi -aqal* 'gather (itr.)'

some doublets, e.g. indigenous *nuzo* (archaic), Avar loan *bet'erhan* 'husband'

phonetic adaptation *r > y* in older loans, e.g. *beyten* 'wedding' < Avar *bertin*

but: *bercinab*, *bak'arzi*, *bet'erhan* – later loans

12. Russian

currently major source of loans into Bezhta (and other Tsezic languages)

including transmission of international vocabulary

even some words with a native counterpart, e.g. *peč* 'stove' (< Russian *peč*) — Bezhta has at least one indigenous word and two loans from

Georgian for different kinds of stoves

also loan adjectives, adverbs, expressions

verbs only in LVCs, in form of infinitive (*-t* < Russian *-t*)

earlier loans have phonetic adaptation

e.g. Russian *fabrika* 'factory' appears as Bezhta *pabrika* and *fabrika*

some early loans transmitted through Georgian have ejectives in Bezhta-Hunzib

e.g. *p'alt'o* < Georgian *p'alt'o* < Russian *pal'to* 'raincoat'

but not Tsez-Hinuq *paltu*

13. Arabic
 Islamicization from C18
 Arabic loans largely restricted to nouns (incl. those used as parts of LVCs)
 also some religious locutions, e.g. *wallah* 'by God'
 one conjunction, *amma* 'but', which is frequent in Bezhta
 semantic domains:
 religion: *allah* 'God', *šäyt'an* 'devil'
 abstract nouns: *hāl* 'state, condition', *pikro* 'thought', *zaman* 'time', *saʔat*
 (also *šäʔät*) 'hour'
 administration: *ahlo* 'people'
 arts: *kalam* 'word', *surat* 'picture'
 science, education: *dars* 'lesson' (often now replaced by Russian, e.g.
 older *müʔälim*, newer *üštel* 'teacher' < *učitel* ')
 names of weekdays, Islamic months, people's names
 stronger phonetic adaptation in Bezhta Proper
 e.g. *hāl* < *ħaal*, *müʔälim* < *muʔallim*
 but without shift *r* > *y*, e.g. *dars* < *dars*

14. Results for chapter 1: The physical world
- | | | |
|-----------------|----|---------------------------------------|
| Indigenous: | 51 | |
| Avar loans: | 9 | of which 2 from Arabic, 1 from Turkic |
| Georgian loans: | 2 | of which 1 questionable |
| Russian: | 3 | of which 1 from Latin, 1 from Greek |
| No equivalent: | 7 | |
| Total | 72 | |

Details:

- | | |
|----------|---|
| Avar | <i>raλ</i> 'land' < <i>raλ</i>
<i>č'ink'ilɬi</i> 'island' < <i>č'ink'ilɬi</i>
<i>räʔäl</i> 'shore' < <i>raʔal</i>
<i>raʔad</i> 'sea' < <i>raʔad</i>
<i>halagab</i> 'rough (of sea)' < <i>halagab</i>
<i>kanɬi</i> 'the light' < <i>kanɬi</i>
<i>dunnal</i> (<i>duniyal</i>) 'the world' < <i>dunyal</i> (<i>duniyal</i>) (< Arabic <i>dunyā</i>)
<i>hawa</i> 'air' < <i>hawa</i> (< Arabic <i>hawā</i>)
<i>maydan</i> 'valley' < <i>maydan</i> 'square, plain' (< Turkic, cf. Kumyk
<i>maydan</i> 'square'); also indigenous <i>ko</i> , less common |
| Georgian | <i>yadri</i> 'embers' < <i>yadari</i>
<i>mašola</i> 'rainbow' < <i>ašuni</i> 'belt'
questionable; Tsez has both meanings 'rainbow', 'belt' |
| Russian | <i>picka</i> 'match' < <i>spička</i>
<i>akean</i> 'ocean' < <i>okean</i> (< Greek <i>ōkeanós</i>)
<i>kantinent</i> 'mainland' < <i>kontinent</i> (< Latin <i>continentum</i>) |

References

- Nikolayev, S.L. and S.A. Starostin. 1994. *A North Caucasian etymological dictionary*. Moscow: Asterisk.
- Xalilov, Madžid Šaripovič. 1993. *Gruzinsko–dagestanskije jazykovye kontakty (na materiale avarsko–cezskix i nekotoryx lezginiskix jazykov)*. Doctoral dissertation, Russian Academy of Sciences, Daghestan Scientific Center, Makhachkala.
- . 1995. *Bežtinsko–russkij slovar*´. Makhachkala: Dagestanskij naučnyj centr Rossijskoj akademii nauk, Institut jazyka, literatury i isskustva im. G. Cadasy.
- . 1999. *Cezsko–russkij slovar*´. Moscow: Academia.
- . 2004. *Gruzinsko–dagestanskije jazykovye kontakty*. Moscow: Nauka.
- and Isaak Abdulvaxidovič Isakov. 2001. *Gunzibsko–russkij slovar*´. Moscow: Nauka.
- and Isaak Abdulvaxidovič Isakov. 2005. *Ginuzsko–russkij slovar*´. Makhachkala: Dagestanskij naučnyj centr Rossijskoj akademii nauk, Institut jazyka, literatury i isskustva im. G. Cadasy.

MEANING		LWT code	1.31	Suggested code		remove meaning	set to an existing meaning >			
view	Semantic Field	PHYSICAL WORLD IN ITS LARGER ASPECTS						Missing word form <input type="checkbox"/>	Language name	Bezhta
	LWT label (M2)	the water						Reason for missing word form		
	Meaning Description (M3)	'drinking water'						Comments (on missing word form)		
	Typical Context (M4)	The inhabitants get their water from the river.								
	Semantic category (M5)	Noun	Next LWT Chapter #		8					
MEANING-WORD RELATIONSHIP										
Word to meaning relationship (W25)										
Comments on meaning-word relationship (W5A)										
FOR ALL WORDS										
view	Record #	10	Default lang	Bezhta	remove word	set to an existing word >				
	Project Language (W0)	Bezhta (Daghestanian)						Original Script (W3)		
	LG	Word Form (W2) ɬi						Comments On Word Form (W5B)		
>>	Meaning (W6)	water						Grammatical Info (W4)		
>>	Analyzability (W7)	Unanalyzable						Calqued (W11)		
	Morpheme By Morpheme Gloss (W8)	Borrowed (W9) 0. No evidence for borrowing						Created On Loan Basis (W12)		
populate age table	Age (W13)							Comments On Borrowed (W10)		
	Start year of period			End year of period			Frequency [numeric] (W13A)	occurrences per million words		
	Record status	INCOMPLETE						Frequency [relative] (W13B)		
								Register (W23)		
ONLY FOR LOANWORDS, CALQUES, AND CREATED ON LOAN BASIS <input type="checkbox"/> No source word (neither W14 nor W14A) is identifiable										
Immediate source word					Earliest known source word					
Word form (W14)					Word form (W14A)					
Donor Language (W15)					Donor Language (W15A)					
Meaning (W16)					Meaning (W16A)					
Effect (W18)					Comments on Intermediate Source Words (W17)					
Environmental Salience (W20)					Integration (W19)					
Reference (W21)					Contact situation (W27)					
Other Comments (W22)										
List of words related to meaning: the water					List of meanings related to word: ɬi					
Word Form (W2)	Word to meaning relationship (W25)				LWT label (M2)					
ɬi				view	the water				view	

MEANING		LWT code: 1.32	Suggested code: <input type="text"/>	remove meaning	set to an existing meaning >
view	Semantic Field: PHYSICAL WORLD IN ITS LARGER ASPECTS	Missing word form <input type="checkbox"/>			
	LWT label (M2): the sea	Language name: Bezhta			
	Meaning Description (M3):	Reason for missing word form: <input type="text"/>			
	Typical Context (M4): The ship is sailing in the sea.	Comments (on missing word form): <input type="text"/>			
	Semantic category (M5): NOUN	Next LWT Chapter #: 8			
MEANING-WORD RELATIONSHIP					
Word to meaning relationship (W25): <input type="text"/>					
Comments on meaning-word relationship (W5A): <input type="text"/>					
FOR ALL WORDS					
view	Record #: 11	Default lang: Bezhta	remove word	set to an existing word >	
	Project Language (W0): Bezhta [Daghestanian]	set default			
LG	Word Form (W2) : raʃad	Original Script (W3):			
<input type="text"/>	Meaning (W6): sea	Comments On Word Form (W5B):			
AUM	Analyzability (W7): Unanalyzable	Grammatical Info (W4):			
<input type="text"/>	Morpheme By Morpheme Gloss (W8):	Calqued (W11):			
	Borrowed (W9) : 4. Clearly borrowed	Created On Loan Basis (W12):			
populate age table	Age (W13):	Comments On Borrowed (W10):			
	Start year of period:	Frequency [numeric] (W13A):	occurrences per million words		
	End year of period:	Frequency [relative] (W13B):			
	Record status: INCOMPLETE	Register (W23):			
ONLY FOR LOANWORDS, CALQUES, AND CREATED ON LOAN BASIS <input type="checkbox"/> No source word (neither W14 nor W14A) is identifiable					
Immediate source word			Earliest known source word		
Word form (W14):	raʃad	Word form (W14A):			
Donor Language (W15):	Avar	Donor Language (W15A):			
Meaning (W16):	sea	Meaning (W16A):			
Effect (W18):	insertion	Comments on Intermediate Source Words (W17):			
Environmental Salience (W20):	Present only since contact	Integration (W19):			
Reference (W21):		Contact situation (W27):			
Other Comments (W22):					
List of words related to meaning: the sea			List of meanings related to word: raʃad		
Word Form (W2)			Word to meaning relationship (W25)		
raʃad <input type="text"/> view			the sea <input type="text"/> view		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		

MEANING		LWT code: 1.841	Suggested code: <input type="text"/>	<input type="button" value="remove meaning"/>	<input type="button" value="set to an existing meaning >"/>	<input type="button" value="view"/>
Semantic Field: PHYSICAL WORLD IN ITS LARGER ASPECTS		LWT label (M2): the embers		Missing word form <input type="checkbox"/> Language name: Bezhta		
Meaning Description (M3): pieces of wood or coal in a fire that are no longer burning but are still red		Typical Context (M4): <input type="text"/>		Reason for missing word form: <input type="text"/>		
Semantic category (M5): Noun		Next LWT Chapter #: 8		Comments (on missing word form): <input type="text"/>		
MEANING-WORD RELATIONSHIP						
Comments on meaning-word relationship (W5A): <input type="text"/>		Word to meaning relationship (W25): <input type="text"/>				
FOR ALL WORDS						
Record #: 72		Default lang: Bezhta		<input type="button" value="remove word"/>		
<input type="button" value="view"/>		<input type="button" value="set default"/>		<input type="button" value="set to an existing word >"/>		
Project Language (W0): Bezhta [Daghestanian]		Original Script (W3): <input type="text"/>		Comments On Word Form (W5B): <input type="text"/>		
LG: Word Form (W2) yadri		Meaning (W6): embers		Grammatical Info (W4): <input type="text"/>		
AUM: <input type="text"/>		Analyzability (W7): Unanalyzable		Calqued (W11): <input type="text"/>		
Morpheme By Morpheme Gloss (W8): Borrowed (W9) 4. Clearly borrowed		Frequency [numeric] (W13A): <input type="text"/>		Created On Loan Basis (W12): <input type="text"/>		
Age (W13): <input type="text"/>		Frequency [relative] (W13B): <input type="text"/>		Comments On Borrowed (W10): <input type="text"/>		
Start year of period: <input type="text"/>		End year of period: <input type="text"/>		occurrences per million words: <input type="text"/>		
Record status: INCOMPLETE		Register (W23): <input type="text"/>				
ONLY FOR LOANWORDS, CALQUES, AND CREATED ON LOAN BASIS <input type="checkbox"/> No source word (neither W14 nor W14A) is identifiable						
		Immediate source word		Earliest known source word		
Word form (W14): yadari		Donor Language (W15): Georgian		Word form (W14A): <input type="text"/>		
Meaning (W16): embers		Effect (W18): No information		Donor Language (W15A): <input type="text"/>		
Environmental Saliency (W20): Present in pre-contact environment		Reference (W21): <input type="text"/>		Meaning (W16A): <input type="text"/>		
Other Comments (W22): <input type="text"/>				Comments on Intermediate Source Words (W17): <input type="text"/>		
				Integration (W19): <input type="text"/>		
				Contact situation (W27): <input type="text"/>		
List of words related to meaning: the embers		List of meanings related to word: yadari				
Word Form (W2)		Word to meaning relationship (W25)				
yadari <input type="button" value="view"/>		LWT label (M2): the embers <input type="button" value="view"/>				
<input type="text"/>		<input type="text"/>				
<input type="text"/>		<input type="text"/>				
<input type="text"/>		<input type="text"/>				
<input type="text"/>		<input type="text"/>				

MEANING		LWT code	1.1	Suggested code		remove meaning	set to an existing meaning >
view	Semantic Field	PHYSICAL WORLD IN ITS LARGER ASPECTS					
	LWT label (M2)	the world					
	Meaning Description (M3)						
	Typical Context (M4)	The Amazon is the longest river in the world.					
	Semantic category (M5)	Noun	Next LWT Chapter #		8		
MEANING-WORD RELATIONSHIP		Word to meaning relationship (W25)					
		Comments on meaning-word relationship (W5A) variant duniyal					
FOR ALL WORDS							
view	Record #	1	Default lang	Bezhta	remove word	set to an existing word >	
	Project Language (W0)	Bezhta [Daghestanian] set default					
	Word Form (W2)	dunnal	variant duniyal				
LG	Meaning (W6)	world					
AUM	Analyzability (W7)	Unanalyzable					
	Morpheme By Morpheme Gloss (W8)	4. Clearly borrowed					
	Borrowed (W9)	4. Clearly borrowed					
populate age table	Age (W13)		Frequency [numeric] (W13A)	occurrences per million words			
	Start year of period		Frequency [relative] (W13B)				
	End year of period		Register (W23)				
	Record status	INCOMPLETE					
ONLY FOR LOANWORDS, CALQUES, AND CREATED ON LOAN BASIS <input type="checkbox"/> No source word (neither W14 nor W14A) is identifiable							
	Immediate source word	Earliest known source word					
	Word form (W14)	dunya, duniyal			dunya		
	Donor Language (W15)	Avar			Arabic		
	Meaning (W16)	world			world		
	Effect (W18)	insertion					
	Environmental Salience (W20)	Present only since contact			1. Highly integrated		
	Reference (W21)						
	Other Comments (W22)						
List of words related to meaning: the world				List of meanings related to word: dunnal			
Word Form (W2)		Word to meaning relationship (W25)		LWT label (M2)			
dunnal			view	the world		view	

MEANING		LWT code: 1.87	Suggested code: <input type="text"/>	remove meaning	set to an existing meaning >
view	Semantic Field: PHYSICAL WORLD IN ITS LARGER ASPECTS	LWT label (M2): the match		<input type="checkbox"/> Missing word form Language name: Bezhta Reason for missing word form: <input type="text"/> Comments (on missing word form): <input type="text"/>	
Meaning Description (M3): a small wooden stick used to light a fire'		Typical Context (M4): <input type="text"/>		Semantic category (M5): Noun	
Next LWT Chapter #: 8					
MEANING-WORD RELATIONSHIP					
Word to meaning relationship (W25): <input type="text"/>					
Comments on meaning-word relationship (W5A): <input type="text"/>					
FOR ALL WORDS					
view	Record #: 49	Default lang: Bezhta	remove word	set to an existing word >	
Project Language (W0): Bezhta [Daghestanian]		set default		Original Script (W3): <input type="text"/>	
LG	Word Form (W2)	picka		Comments On Word Form (W5B): <input type="text"/>	
<input type="text"/>	Meaning (W6)	match		Grammatical Info (W4): <input type="text"/>	
AUM	Analyzability (W7)	Unanalyzable		Calqued (W11): <input type="text"/>	
<input type="text"/>	Morpheme By Morpheme Gloss (W8)	Borrowed (W9): 4. Clearly borrowed		Created On Loan Basis (W12): <input type="text"/>	
Age (W13): <input type="text"/>		Frequency [numeric] (W13A): <input type="text"/>		Comments On Borrowed (W10): <input type="text"/>	
Start year of period: <input type="text"/>		End year of period: <input type="text"/>		Frequency [relative] (W13B): <input type="text"/>	
Record status: INCOMPLETE		occurrences per million words: <input type="text"/>		Register (W23): <input type="text"/>	
ONLY FOR LOANWORDS, CALQUES, AND CREATED ON LOAN BASIS <input type="checkbox"/> No source word (neither W14 nor W14A) is identifiable					
Immediate source word			Earliest known source word		
Word form (W14):	spička		Word form (W14A):	<input type="text"/>	
Donor Language (W15):	Russian		Donor Language (W15A):	<input type="text"/>	
Meaning (W16):	match		Meaning (W16A):	<input type="text"/>	
Effect (W18):	<input type="text"/>		Comments on Intermediate Source Words (W17):	<input type="text"/>	
Environmental Salience (W20):	<input type="text"/>		Integration (W19):	<input type="text"/>	
Reference (W21):	<input type="text"/>		Contact situation (W27):	<input type="text"/>	
Other Comments (W22):	<input type="text"/>				
List of words related to meaning: the match			List of meanings related to word: picka		
Word Form (W2)			Word to meaning relationship (W25)		
picka <input type="text"/> view			LWT label (M2)		
<input type="text"/>			the match <input type="text"/> view		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		
<input type="text"/>			<input type="text"/>		

