

Language Specific Peculiarities Document for Georgian as Spoken in Georgia

Georgian is a Kartvelian language, spoken in Georgia in the Caucasus. It is spoken by approximately 4 million speakers.

1. Special handling of dialects

Georgian is commonly split into two major regional dialect groups – eastern and western (ARMAZI, 2002). These groups can be further divided into 14 separate dialects spoken in Georgia as shown in the table below, although the distinction between them is often restricted to lexical choices (Gigineishvili et al., 1961).

Region	Dialects
Eastern	Javakhian, Kartlian, Meskhian, Kakhetian, Khevsurian, Mokhevian, Kizlar-Mozdokian, Pshavian, Tushetian
Western	Imeretian, Lechkhumian, Rachan, Adjarian, Gurian

Gigineishvili et al. (1961) uses five dialect groups instead of the broader two-group split. This grouping, however, is not taken up in most of the more recent literature (and the two group split was confirmed by native-speaker linguist research).

The standard dialect is Kartlian, spoken in the region of Kartli. The majority of speakers in the capital city Tbilisi also speak Kartlian Georgian, although there is a great deal of dialectal variety in Tbilisi.

Most dialects spoken in Georgia are mutually intelligible between speakers. Some have higher degrees of lexical variation and therefore are more difficult for speakers of standard Georgian to understand; Gurian and Khevsurian use such different lexemes as to approach unintelligibility (Jorbenadze, 1989).

Speakers from both dialect groups above have been included in this speech database. It is important to note, however, that the speech database may not cover all sub-dialects as a number of these sub-dialects are spoken by very small communities, possibly of as little as 3,000 people.

Three dialects of Georgian are not spoken in Georgia: Fereydanian, Imerkhavian, and Ingolian. These are spoken natively in Iran, Turkey, and Azerbaijan respectively, and were not collected for the current project.

There are also three related Kartvelian languages spoken in Georgia: Laz, Megrelian, and Svan. These languages are not mutually intelligible with Georgian, and they were not collected for this project.

2. Deviation from native-speaker principle

There is no deviation from the native-speaker principle for this project. Only native speakers of Georgian in Georgia were recruited.

3. Special handling of spelling

None.

4. Description of character set used for orthographic transcription

Georgian has its own alphabet. The script used to write modern standard Georgian is Mkhedruli. This alphabet is found in the Unicode range 10D0-10F0. See the table in section 5 for a mapping of each character to its code point.

5. Description of Romanization scheme

The table below contains Appen's Romanization scheme, which is fully reversible. Transcription work was done by Georgian speakers working with the Mkhedruli alphabet and no Romanization. The Romanization scheme was primarily used as a reference for those unfamiliar with the Georgian script.

The Appen Romanization scheme is similar to most common transliterations used when writing Georgian in Latin script. The major change is in the use of capital letters, rather than diacritics or digraphs to represent the fricatives and affricates: /j/ as <S>, /z/ as <Z>, /x/ as <K>, /ɣ/ as <G>, /ts/ as <T>, /dz/ as <D>, and /tʃ/ as <C>. The practice of representing some of these sounds as digraphs was not reversible, since <h>, <s>, and <z> are already used to represent independent sounds.

In this Romanization, the apostrophe <'> is used to represent glottalization, not aspiration. Ejective consonants are represented by placing an apostrophe <'> after the symbol for the corresponding non-ejective plosive or affricate, for example, /t'/ as <t'>, /ts'/ as <T'>. Note that the uvular ejective plosive <q'> does not have a non-ejective counterpart, since in the synchronic language /q/ and /q'/ have merged to /q'/.

UNICODE	GEORGIAN	ROMAN	DESCRIPTION
0x10D0	ა	a	GEORGIAN LETTER AN
0x10D1	ბ	b	GEORGIAN LETTER BAN
0x10D2	გ	g	GEORGIAN LETTER GAN
0x10D3	დ	d	GEORGIAN LETTER DON
0x10D4	ე	e	GEORGIAN LETTER EN
0x10D5	ვ	v	GEORGIAN LETTER VIN
0x10D6	ზ	z	GEORGIAN LETTER ZEN

UNICODE	GEORGIAN	ROMAN	DESCRIPTION
0x10D7	თ	t	GEORGIAN LETTER TAN
0x10D8	ო	i	GEORGIAN LETTER IN
0x10D9	კ	k'	GEORGIAN LETTER KAN
0x10DA	ლ	l	GEORGIAN LETTER LAS
0x10DB	მ	m	GEORGIAN LETTER MAN
0x10DC	ნ	n	GEORGIAN LETTER NAR
0x10DD	ო	o	GEORGIAN LETTER ON
0x10DE	პ	p'	GEORGIAN LETTER PAR
0x10DF	ჯ	Z	GEORGIAN LETTER ZHAR
0x10E0	რ	r	GEORGIAN LETTER RAE
0x10E1	ს	s	GEORGIAN LETTER SAN
0x10E2	ტ	t'	GEORGIAN LETTER TAR
0x10E3	უ	u	GEORGIAN LETTER UN
0x10E4	ფ	p	GEORGIAN LETTER PHAR
0x10E5	ქ	k	GEORGIAN LETTER KHAR
0x10E6	ღ	G	GEORGIAN LETTER GHAN
0x10E7	ყ	q'	GEORGIAN LETTER QAR
0x10E8	შ	S	GEORGIAN LETTER SHIN
0x10E9	ჩ	C	GEORGIAN LETTER CHIN
0x10EA	ც	T	GEORGIAN LETTER CAN
0x10EB	ძ	D	GEORGIAN LETTER JIL
0x10EC	წ	T'	GEORGIAN LETTER CIL
0x10ED	ჭ	C'	GEORGIAN LETTER CHAR
0x10EE	ხ	K	GEORGIAN LETTER XAN
0x10EF	ჯ	j	GEORGIAN LETTER JHAN
0x10F0	ჰ	h	GEORGIAN LETTER HAE

6. Description of method for word boundary detection

Word boundaries in the orthography are determined by whitespace.

Georgian is an agglutinative language – words can be given multiple suffixes. These suffixes are always joined to the stem, with no intervening whitespace.

Some acronyms exist in Georgian, but these are spelled without the intervening periods which are sometimes used in English. If a sequence is read out letter by letter (which may be the case for some business or administrative names), the letters have been spelled separated by underscores, according to the usual transcription conventions for this project.

7. Table containing all phonemes in the stipulated notation

The phonemic transcription of the words in this database uses X-SAMPA symbols, which can be found at <http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>. The total number of native phonemes is 33. There are 28 consonants and 5 vowels (all monophthongs) (Aronson, 1990).

GEORGIAN PHONE CHART

<i>TYPICAL GEORGIAN CORRESPONDENCE</i>	UNICODE	ROMAN	IPA	SAMPA	COMMENTS
CONSONANTS					
ბ	0x10D1	b	b	b	Realized as [p_h] in some contexts
გ	0x10D2	g	g	g	Realized as [k_h] in some contexts
დ	0x10D3	d	d	d	Realized as [t_h] in some contexts
ვ	0x10D5	v	v	v	Realized closer to [w] by some speakers
ზ	0x10D6	z	z	z	
თ	0x10D7	t	t ^h	t_h	
კ	0x10D9	k'	k'	k_>	
ლ	0x10DA	l	l	l	
მ	0x10DB	m	m	m	

TYPICAL GEORGIAN CORRESPONDENCE	UNICODE	ROMAN	IPA	SAMPA	COMMENTS
ნ	0x10DC	n	n	n	
პ	0x10DE	p'	p'	p_>	
ჯ	0x10DF	Z	ʒ	Z	Rare phone
რ	0x10E0	r	r	r	
ს	0x10E1	s	s	s	
ტ	0x10E2	t'	t'	t_>	
			p ^h	p_h	
ფ	0x10E4	p	f	f	Foreign phone
კ	0x10E5	k	k ^h	k_h	
გ	0x10E6	G	ɣ	G	
ყ	0x10E7	q'	q'	q_>	
შ	0x10E8	S	ʃ	S	
ჩ	0x10E9	C	tʃ ^h	tS_h	
ც	0x10EA	T	ts ^h	ts_h	
ძ	0x10EB	D	dʒ	dz	
წ	0x10EC	T'	ts'	ts_>	
ჭ	0x10ED	C'	tʃ'	tS_>	
ხ	0x10EE	K	x	x	
ჯ	0x10EF	j	dʒ	dZ	
ჰ	0x10F0	h	h	h	Rare phone
VOWELS					

TYPICAL GEORGIAN CORRESPONDENCE	UNICODE	ROMAN	IPA	SAMPA	COMMENTS
ა	0x10D0	a	a	a	
ე	0x10D4	e	ɛ	E	
ი	0x10D8	i	i	i	ი is also used to represent this foreign semi-vowel. Rare.
			j	j	
ო	0x10DD	o	ɔ	O	
უ	0x10E3	u	u	u	

OTHER SYMBOLS	
.	syllable break
#	word boundary

Notes

1. Stress is very weak in Georgian, and is not used for phonemic contrast (Aronson, 1990). It has not been marked for this project.
2. The primary pronunciation given in the lexicon for loan words reflects careful speech by an educated speaker, rather than following the Georgian orthography of the word. A nativized pronunciation has been given as a variant pronunciation in these cases.

7.1 List of rare phones

There are of course varying frequencies of use between the various phonemes. The phones /z/ and /h/ and the foreign phones listed in section 7.2 occur less frequently than others in spoken Georgian.

7.2 List of foreign phones

The phones /f/ and /j/ are commonly used in loanwords – the latter in particular in nativizing rising diphthongs, e.g. /ju/ and /ja/. Since the Georgian alphabet has no provisions to represent these phones, they are nativized using the symbol <ფ> for /f/ and the symbol <ი> for /j/.

8. Other language items

8.1 Table of Digits

Numeral	Georgian	Romanized
0	ნული	nuli
1	ერთი	erti
2	ორი	ori
3	სამი	sami
4	ოთხი	otKi
5	ხუთი	Kuti
6	ექვსი	ekvsi
7	შვიდი	Svidi
8	რვა	rva
9	ცხრა	TKra

8.2 Other Numbers

Numeral	Georgian	Romanized
10	ათი	ati
100	ასი	asi
10,000	ათი ათასი	ati atasi
100,000	ასი ათასი	asi atasi
10 million	მილიონი	milioni

8.3 Spelling Alphabet

Letters in the Georgian alphabet have formal names (Hewitt, 1996). The character ა (<a>) for example, would be named /a . n i/ in citation form, or /a n/ in inflected form.

However, these names are not typically used in spelled items in Georgian. The pronunciation for spelled items is typically the simple vowel sound for vowels, and the consonant followed by a short vowel for consonants, as represented below.

Orthography	Usual pronunciation (SAMPA)
ა	a

Orthography	Usual pronunciation (SAMPA)
ɸ	b E
ɸ	g E
ʃ	d E
ɔ	E
ʒ	v E
ʃ	z E
ʈ	t_h E
o	i
ʒ	k_> E
ʃ	l E
ɸ	m E
ʃ	n E
ʒ	o
ʒ	p_> E
ʃ	z E
ʃ	r E
ʃ	s E
ʃ	t_> E
ʃ	u
ʃ	p_h E
ʃ	k_h E
ʃ	g E
ʃ	q_> E
ʃ	s E
ʃ	tS_h E
ʃ	ts_h E
ɸ	dz E

Orthography	Usual pronunciation (SAMPA)
ჭ	ts_> E
ჭ	tS_> E
ბ	x E
ჯ	dZ E
ჰ	h E

9. References

- ARMAZI. 2002. http://www.armazi.demon.co.uk/georgian/unicode/georgian_dialects.htm, retrieved May 14th, 2014.
- Aronson, H. I. 1990. *Georgian: A reading grammar*, Slavica. Bloomington, IN.
- Awde, N. and Khitarishvili, T. 1997. *Georgian. Hippocrene Dictionary and Phrasebook*. Hippocrene Books, New York.
- Beridze, M., Lortkipanidze, L., and Nadaraia, D. 2013. Dialect Dictionaries in the Georgian Dialect Corpus, in *Logic, Language, and Computation*, Martin Aher, Daniel Hole, Emil Jerabek, and Clemens Kupke (eds.). Springer, Heidelberg, pp. 82-96.
- Butskhrikidze, M. 2002. *The consonant phonotactics of Georgian*. LOT, Utrecht.
- Gigineishvili, I., Topuria, V., and Kavtaradze, I. 1961. *kartuli dialekt'ologia I* (Georgian dialectology, I), TSUG.
- Gillam, R. 2002. *Unicode Demystified: A Practical Programmer's Guide to the Encoding Standard*, Addison-Wesley Professional.
- Georgian dialect corpus. n.d¹. <http://www.mygeorgia.ge/gdc/About.aspx>, retrieved May 14th, 2011.
- Harrell, S. J., Koupounla, M., Tsitsishvili, M., Gabounia, K. 2002. *Georgian-English Dictionary*. Dunwoody Press, Springfield, VA.
- Hewitt, G. 1996. *Georgian: A learner's grammar*. Routledge, London and New York.
- Holisky, D. A. 1996. The Georgian alphabet, in *The World's Writing Systems*, Peter T. Daniels and William Bright (eds.), Oxford & New York: *Oxford University Press*, pp. 364-369.
- Jorbenadze, B. 1989. *kartuli dialekt'ologia I* (Georgian Dialectology, I). Tbilisi, Mecniereba.
- Kojima, Y. 2007. Clitics in modern Georgian. *Asian and African Languages and Linguistics* 2. pp. 85-102.
- Robins, R. H. and Waterson, N. 1952. *Notes on the Phonetics of the Georgian Word*, in *Bulletin*

¹ n.d. means no date.

of the School of Oriental and African Studies 14.

Tuite, K. 1998. *Kartvelian Morphosyntax: Number Agreement and Morphosyntactic Orientation in the South Caucasian Languages*. LINCOS Studies in Caucasian Linguistics 12. Lincom Europa, Munich.