

Language Specific Peculiarities Document for

Lithuanian as Spoken in Lithuania

1. Dialects

Lithuanian, spoken in Lithuania, is one of three languages belonging to the Eastern subgroup of the Baltic language family. The other two languages belonging to this subgroup are Latvian and Latgalian which are both spoken in Lithuania’s neighbouring country, Latvia (Ethnologue, 2013).

There are two main regional dialects of Lithuanian – Aukštaitian (a.k.a. High Lithuanian) and Samogitian (a.k.a. Žemaitian or Low Lithuanian). Standard Lithuanian is based on the Aukštaitian dialect. This dialect is spoken in approximately three-quarters of the country. People have traditionally only spoken Samogitian dialects in the north-west.

Table 1. Lithuanian Dialect Distribution (Girdenio & Zinkevičiaus 1977 as cited by Lietuvių Kalbos Institutas)

Dialect name	Districts or Cities
Aukštaitian (High Lithuanian)	Vilnius, Ukmergė, Panevėžys, Pakruojis, Biržai, Širvintos, Anykščiai, Kupiškis, Rokiškis, Utena, Molėtai, Ignalina, Jurbarkas, Kaunas, Prienai, Kaišiadorys, Jonava, Kėdainiai, Joniškis, Šiauliai, Lazdijai, Alytus, Šalčinikai
Samogitian or Žemaitian (Low Lithuanian)	Šilutė, Klaipėda, Kretinga, Skuodas, Telšiai, Najoji Akmenė, Varniai, Raseiniai, Tauragė

Mutual Intelligibility of Standard Lithuanian and Samogitian

Most ethnic Lithuanians, and many members of ethnic minorities in Lithuania are competent in standard Lithuanian, the language variety that is codified in grammars and taught in the classroom. Speakers of standard Lithuanian may find some regional dialects, such as Samogitian, difficult to understand; however, speakers of all regional dialects generally understand standard Lithuanian.

2. Deviation from native-speaker principle

No special deviation – only data from native speakers of Lithuanian, born in Lithuania, will be collected in this project.

3. Special handling of spelling

Lithuanian spelling was standardised in late 19th century and has not changed since that time, save for a few minutiae. There are no alternative spelling systems of Lithuanian to be taken into

account for the purpose of language processing. The standardised orthography is used for all writing. The characters needed for the standard orthography are listed in Table 2, Section 4 below.

3.1 Standard dictionaries

Two reliable sources for Lithuanian spelling are the following dictionaries: *Lietuvių kalbos žodynas* (*Dictionary of Lithuanian Language*, LKŽ) and *Dabartinės lietuvių kalbos žodynas* (*Dictionary of Contemporary Lithuanian Language*, DLKŽ). In the very rare cases when spellings in those two dictionaries differ, DLKŽ is to be preferred; but LKŽ is much bigger, and includes those rare (dialectal, obsolete etc.) words which are absent from DLKŽ. Both of those dictionaries are freely available online.

- *Lietuvių kalbos žodynas* (*Dictionary of Lithuanian Language*):
<http://www.lkz.lt/startas.htm>
- *Dabartinės lietuvių kalbos žodynas* (*Dictionary of Contemporary Lithuanian Language*):
<http://dz.lki.lt/>

Note: the headwords are spelled with pitch marking in these dictionaries. In order to extract the standard spelling, pitch marks (acute, grave, tilde-shaped circumflex) must be removed from the headwords.

4. Description of character set used for orthographic transcription

Lithuanian uses the Basic Latin Unicode range (U+0041-U+007A) plus the following additional forms from the Latin Extended-A set.

Symbol	Unicode	Description
Ą	U+0104	Latin Capital Letter A with ogonek
ą	U+0105	Latin Small Letter A with ogonek
Č	U+010C	Latin Capital Letter C with caron
č	U+010D	Latin Small Letter C with caron
Ę	U+0118	Latin Capital Letter E with ogonek
ę	U+0119	Latin Small Letter E with ogonek
Ė	U+0116	Latin Capital Letter E with dot above
ė	U+0117	Latin Small Letter E with dot above
Į	U+012E	Latin Capital Letter I with ogonek
į	U+012F	Latin Small Letter I with ogonek

Symbol	Unicode	Description
Ą	U+0104	Latin Capital Letter A with ogonek
ą	U+0105	Latin Small Letter A with ogonek
Š	U+0160	Latin Capital Letter S with caron
š	U+0161	Latin Small Letter S with caron
Ų	U+0172	Latin Capital Letter U with ogonek
ų	U+0173	Latin Small Letter U with ogonek
Ū	U+016A	Latin Capital Letter U with macron
ū	U+016B	Latin Small Letter U with macron
Ž	U+017D	Latin Capital Letter Z with caron
ž	U+017E	Latin Small Letter Z with caron

5. Description of Romanization scheme

Not applicable.

6. Description of method for word boundary detection

Word boundaries in the orthography are determined by white space.

Hyphens are rarely used in Lithuanian. Except for very few compounds which are spelt with a hyphen, hyphens are used only to separate double surnames.

Acronyms in Lithuanian are typically written in upper case, and the letters are not separated by full stops, e.g. JAV (Jungtinės Amerikos Valstijos, i.e. United States of America).

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 phonemes is 56 and there are 2 phonological tones. The set of phonemes comprises 45 consonants and 11 vowels.

It should be noted that Lithuanian also has two diphthongs (namely ie and uo) and seven vowel sequences (ui, ei, ai, au, eu, oi, ou) (see Ambrasas 2006). We propose to treat both vowel sequences and diphthongs as vowel sequences for the purposes of this project as there do not appear to be significant phonological reasons to treat them differently. The reason they are sometimes treated distinctly relates principally to sub-phonemic differences in vowel lengthening under stress in vowel sequences (which are not observed in the sequences referred

to as diphthongs). Relatively longer vowel length is also observed in single vowels compared with vowels in vowel sequences.

Tone distinction is only possible in long or ‘heavy’ syllables. In traditional school practice, the short syllable stress was also considered a ‘tone’, but will not be considered as such for our purpose. Long or ‘heavy’ syllables take either falling (acute) or rising (circumflex) tone. These syllables have either a long vowel, a diphthong, a vowel sequence or a ‘mixed diphthong’ combining a short vowel and a sonorant as the syllable nucleus. See the ‘Tone/Pitch Accent’ table below for tone/pitch accent symbols.

LITHUANIAN PHONE CHART
(Press 2004, Kenstowicz 1972)

TYPICAL LITHUANIAN CORRESPONDENCE	IPA	SAMPA	COMMENTS
CONSONANTS			
p	p p ^j	p p'	non-palatalized palatalized
b	b b ^j	b b'	non-palatalized palatalized
t	t t ^j	t t'	non-palatalized palatalized
d	d d ^j	d d'	non-palatalized palatalized
k	k k ^j	k k'	non-palatalized palatalized
g	g g ^j	g g'	non-palatalized palatalized
v	v v ^j	v v'	non-palatalized palatalized
s	s s ^j	s s'	non-palatalized palatalized
z	z z ^j	z z'	non-palatalized palatalized
š	ʃ ʃ ^j	S S'	non-palatalized palatalized
ž	ʒ ʒ ^j	Z Z'	non-palatalized palatalized

TYPICAL LITHUANIAN CORRESPONDENCE	IPA	SAMPA	COMMENTS
c	t tʲ	ts ts'	non-palatalized palatalized
dz	dʒ dʒʲ	dz dz'	non-palatalized - rare palatalized - rare
č	tʃ tʃʲ	tʃ tʃ'	non-palatalized - rare palatalized
dž	dʒ dʒʲ	dʒ dʒ'	non-palatalized - rare palatalized
m	m mʲ	m m'	non-palatalized palatalized
n	n nʲ	n n'	non-palatalized palatalized
l	l lʲ	l l'	non-palatalized palatalized
r	r rʲ	r r'	non-palatalized palatalized
j	j	j	always palatal
f	f fʲ	f f'	used only in foreign borrowings and some interjections.
ch	x xʲ	x x'	
h	ɣ ɣʲ	G G'	
VOWELS			
a	a	a	
a, ą	ɑ:	A:	
e	ɛ	E	
e, ę	æ:	{:	
ė	e:	e:	
į, y	i:	i:	

TYPICAL LITHUANIAN CORRESPONDENCE	IPA	SAMPA	COMMENTS
i	ɪ	ɪ	
o	o:	o:	
o	ɔ	ɒ	used only in foreign borrowings
u	ʊ	ʊ	
ų, ū	u:	u:	
e (after i)	e	e	only used in vowel sequences
i (after u)	i	i	only used in vowel sequences

The following symbols occur in the lexicon attached to the appropriate vowel or sonorant, e.g. /A:_F/.

TONE/PITCH ACCENT	
_R	Rising contour
_F	Falling contour

In the case of diphthongs (two vowels) and “semi” or “mixed” diphthongs (a vowel followed by a sonorant /r/, /l/, /m/, or /n/) falling contour is usually marked on the initial vowel of the sequence, while rising contour is usually marked on the final vowel or sonorant. In some pronunciation variants, rising contour may also occur on the initial vowel of the sequence, and is marked as such in the transcriptions.

OTHER SYMBOLS	
.	syllable break
#	word boundary

7.1 List of rare phonemes

The following phonemes are rare (already listed in the table above):

TYPICAL LITHUANIAN CORRESPONDENCE	IPA	SAMPA	COMMENTS
č	tʃ	tS	Used only in loanword and some interjections.
dž	dʒ	dZ	
f	f fʲ	f f'	
ch	x xʲ	x x'	
h	ɣ ɣʲ	G G'	
c	ts tsʲ	ts ts'	used mostly in loanwords and onomatopoeic words
dz	dʒ dʒʲ	dz dz'	
e (after i)	e	e	only used in vowel sequences

General note:

As some of these consonant phonemes occur in relatively recent loanwords, they may be rare among types; however, since some of those loanwords are very frequently used, they may not be rare among tokens. For example:

- Unpalatalized /tʃ/ is very rare, restricted mostly to loanwords and expressive vocabulary except for *ginčas* 'argument'; *ypač* 'especially'.
- /f, f'/ are found in frequent, well-established words such as *filmas* 'film,' *firma* 'firm,' and *futbolas* 'football, soccer.' Not necessarily "rare" in the synchronic phonology (*fonologija*), however, palatalized /f/ is rare (except *amfibrachis* 'amphibrach').
- /x, x'/ are found in a few frequent, well-established words such as *chaosas* 'chaos,' *choras*, 'chorus,' *chuliganas* 'hooligan'; however, palatalized /x/ is rare.
/G, G'/ are found in a few frequent, well-established words such as *humoras* 'humor,' *himnas* 'hymn, anthem,' *herojus* 'hero'.

7.2 List of foreign phones

The following phonemes are foreign (already listed in the table above):

TYPICAL LITHUANIAN CORRESPONDENCE	IPA	SAMPA	COMMENTS
f	f fʲ	f f'	Used only in foreign borrowings and some interjections.
ch	x xʲ	x x'	
h	ɣ ɣʲ	G G'	

8. Other Language Specific Items

8.1 Table of Digits

Digit	Word
0	nulis
1	vienas
2	du
3	trys
4	keturi
5	penki
6	šeši
7	septyni
8	aštuoni
9	devyni

Note: all digit names are declined; depending on context, they appear in the appropriate case, for example: *penki litai* ('5 Lt', nominative), *penkis litus* (accusative), etc.

All digit names except *nulis* (0), which is masculine (and grammatically more of a noun than a numeral), and *trys* (3), whose masculine and feminine genders coincide, can be either of masculine or feminine gender, for example: *du vyrai* (two men); *dvi moterys* (two women).

In the above table, all digits are given in citation form, which is identical with masculine gender, nominative case. This citation form is expected when dictating telephone numbers, etc.

8.2 Other Numbers

Digit	Word
10	dešimt
100	šimtas
10,000	dešimt tūkstančių
100,000	šimtas tūkstančių
10,000,000	dešimt milijonų

Note: the numeral *dešimt* (10) is indeclinable, but when the case needs to be clarified or emphasised in some contexts, the appropriate case of *dešimtis* ('a group of ten', noun) is used. The numeral *šimtas* (100) is declined as a masculine noun.

Hundreds are counted to 900; in quantities more than 1000, thousands rather than hundreds are counted. That is, 1500 in Lithuanian is *tūkstantis penki šimtai*, literally 'thousand five hundreds', not 'fifteen hundred'.

9. References

Ambrasas, Vytautas et al. (2006), *Dabartinės lietuvių kalbos gramatika (A Grammar of Modern Lithuanian)*, Vilnius: MELI

Dambriunas, L., A. Klimas, et al. (1998). *Beginner's Lithuanian*, New York: Hippocrene Books.

Kenstowicz, Michael (1972), Lithuanian phonology, *Studies in the Linguistic Sciences* 2(2): Illinois. Retrieved from

https://www.ideals.illinois.edu/bitstream/handle/2142/26464/SLS1972v2-2_kenstowicz.pdf?sequence=2 (25th July 2013)

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.) (2013), *Ethnologue: Languages of the World*, Seventeenth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>

Lietuvių kalbos institutas, (2013), Tarmių skirstymas, Retrieved from http://www.tarmes.lt/index_meniu.php?id=1 (25th July 2013)

Mathiassen, Terje. 1996. *A short grammar of Lithuanian*. Columbus, OH: Slavica

Map of dialects of Lithuanian Language (2010), Retrieved from https://commons.wikimedia.org/wiki/File:Map_of_dialects_of_Lithuanian_language.png (25th July 2013)

Piesarskas, Bronius & Bronius Svecevičius. (1995), *Lithuanian dictionary*, London & New York: Routledge.

Press, Ian (2004), *Standard Lithuanian*, Munich: LINCOM

Schmalstieg, William (1982), Standard Lithuanian and its dialects, *Lituanus* 28(1), 27-44.

Retrieved from http://www.lituanus.org/1982_1/82_1_02.htm (25th July 2013)

Standard dictionaries:

Lietuvių kalbos žodynas (Dictionary of Lithuanian Language): <http://www.lkz.lt/startas.htm>

Dabartinės lietuvių kalbos žodynas (Dictionary of Contemporary Lithuanian Language):
<http://dz.lki.lt/>