The Root Structure of Lithuanian Inflective Words

Lietuvių kalbos kaitomųjų žodžių šaknies struktūra

SAL 33/2018

The Root Structure of Lithuanian Inflective Words

Received 02/2018

Accepted 10/2018

LINGUISTICS / KALBOTYRA

Asta Kazlauskienė

Department of Lithuanian Studies, Vytautas Magnus University, Lithuania

Jurgita Cvilikaitė

Institute of Foreign Languages, Vytautas Magnus University, Lithuania

http://dx.doi.org/10.5755/j01.sal.0.33.20179

The aim of this research is to identify the structural patterns of root morphemes of Lithuanian inflective words and to establish their productivity and frequency. First, with reference to the earlier work conducted by Lithuanian linguists, we discuss the structural diversity of root morphemes and determine the productivity of structural patterns (the number of different roots of a specific pattern). Then we analyse data from real usage. For this stage, the database of the morphemics of the Lithuanian language (Lietuvių kalbos morfemikos duomenų bazė) was used. 265 thousand usage instances of inflective words constitute the research data.

The analysis of the root structure allows drawing the following conclusions: 1) although the diversity of morpheme structure is rich, only roots of simple structure are productive and frequent (roots whose onsets or codas contain one to two consonants), 2) root morphemes are non-syllabic or vary from monosyllabic to trisyllabic (non-syllabic roots are the most productive and the most frequent), 3) consonant clusters are not frequent in the middle of a morpheme (they were identified in a third of all roots), 4) the number of consonants in a root usually does not exceed six, 5) onset consonant clusters concur with the pattern of a syllable beginning; consonant clusters in codas and medial clusters are more diverse.

KEYWORDS: morpheme, root, vowel, consonant, inflective part of speech.

Introduction

Linguists have been interested in morpheme structure for a long time. Their interest was induced by the article by Trubetzkoy (1931), who set tasks of morphonology, including the study of the phonological structure of morphemes. The other stimulus came from Jakobson's (1962) idea that different grammatical classes can be characterized by different usage of phonemes.

The aim of the research presented here is to identify the main structural patterns of root morphemes of Lithuanian inflective words (nouns, adjectives, pronouns, numerals and verbs), their productivity and frequency. The research of such kind was motivated by the experience of working with abundant usage and dictionary data, which has repeatedly shown that usage can disclose interesting cases resistant to any prejudice.

Research Journal
Studies about Languages
No. 33/2018
ISSN 1648-2824 (print)
ISSN 2029-7203 (online)
pp. 5-16
DOI 10.5755/j01.sal.33.0.20179

The root structure of verbs provided in dictionaries has been analysed in works of Kruopienė (2000); Kaukienė (1994, 2002) has studied the morphonological structure of verb roots; Akelaitienė (1996, 2000) has been interested in vowel change for many years; the structure of nominal words has been analysed by Karosienė (2004). The aforementioned researchers analyse dictionary but not usage data. Therefore, based on such data we can say which patterns of morphemic structure or sound change are possible in the Lithuanian language; we can also indicate the productivity of a particular pattern; however, we cannot say anything about their frequency. In this article, productivity is perceived as realization of a pattern which can be calculated based on dictionaries. Frequency shows usage instances in real connected speech.

The research and results will be presented in the following stages: first, we provide a survey of the results of research conducted by Lithuanian linguists. Then we analyse and discuss the results of real usage (the frequency of structural patterns is identified). The frequency data for the study was collected from the database of the morphemics of the Lithuanian language (*Lietuvių kalbos morfemikos duomenų bazė*, further in the text DbML), which was created at Vytautas Magnus University and served as the basis for morphemics dictionaries (see Rimkutė, Kazlauskienė, Raškinis 2011). The database contains approximately 72 thousand words from different text styles and various topics. For the extraction of the empirical data, a special computer programme was created[1]. The programme works in several stages: 1) a list (dictionary) of different word forms used in the research material was generated and usage number (frequency) of each word form was identified; 2) words were automatically stressed (in Lithuanian the stress is free) and transcribed with tools available to us (for the latter, see Kazlauskienė, Raškinis, Vaičiūnas 2010), mistakes were corrected manually; 3) morpheme boundaries in all words were marked by conventional symbols manually; 4) words were replaced by conventional symbols[2]: C - for consonants, V - for vowels, W - for diphthongs *ai, au, ei, ui* [ɐɪ, ɐʊ, ɛɪ, ʊɪ]; 5) the received word code CVCW... was segmented into morphemes according to the labels used in words to mark morpheme boundaries; 6) morphemes matching the same code were grouped and their usage frequency in the research data was calculated. The research data from DbML encompasses 174,200 nominal words (103,681 nouns, 23,714 adjectives, 41,139 pronouns, and 5,666 numerals) and 90,800 verbs.

Non-inflective words have been left outside the scope of the study because quite a large part thereof are primary (only a root morpheme, e.g., *ir* 'and', *ar* 'whether', *o* 'and/while', *ne* 'no'); moreover, some of them are grammatical forms of multiword expressions (e.g., *iš tiesų* 'truly', cf. N Pl Gen). Proper nouns were not analysed either, because their morphemic segmentation is complicated.

International words have not been analysed either, as they usually came to the language as morphemically unsegmented words showing structural regularities of other languages rather than those of Lithuanian. However, the old borrowings, e.g., *stiklas 'glass', agurkas* 'a cucumber', have been included into our research as they have adapted in the language, adjusted to the system of the Lithuanian language, and most Lithuanians do not identify them as borrowings.

1 The author of the programme is prof. dr. Gailius Raškinis. We are thankful for his help during the extraction of the empirical data.

2 There are 11 vowels in the Lithuanian language: long [æː, eː, iː, uː, oː, ɑː, ɪɛ, ʊɔ] and short [ɛ, ɪ, ʊ, ɔ, ɐ]. Consonants, depending on their position in a consonant cluster and articulation which is the focus of attention in the analysis of structural types of morphemes, are classified into three groups: T group (plosive consonants [b, bʲ, d, dʲ, g, gʲ, k, kʲ, p, pʲ, t, tʲ] and affricates [t͡s, t͡sʲ, d͡z, d͡zʲ, t͡ʃ, t͡ʃʲ, d͡ʒ, d͡ʒʲ]), S (fricative consonants [s, sʲ, z, zʲ, ʃ, ʃʲ, ʒ, ʒʲ, x, xʲ, f, fʲ, ɣ, ɣʲ]) and R (nasals [m, mʲ, n, nʲ, ŋ, ŋʲ], vibrants [r, rʲ], approximants [j, ʋ, ʋʲ] and lateral approximants [l, lʲ]). The Lithuanian language has diphthongs which form the nucleus of a syllable: [ɐɪ, ɐʊ, ɛɪ, ʊɪ] (see more about the phonological system of the Lithuanian language in Girdenis 2014).

In inflectional languages like Lithuanian, the boundary between synchrony and diachrony is not always clear. Quite often fusion occurs, i.e., morphemes merge; for this reason, affixes are often considered as part of a root. For example, from the modern language perspective, *gyv-* can be considered as the root of the adjective *gyvas* 'alive', although historically, this adjective should be linked to the verb *gyti* 'to recover'. In this case, the morphemic division is between the root *gy-*, the suffix *-v-* and the flexion *-as*. The other example is the adjective *kuklus* 'modest' that is historically related to *kukti* 'to bow'; however, the adjective has digressed from its underlying word considerably in respect of its meaning. During the preparation of the data for the research, we tried to follow the synchronic principle. If the link of a word to a possible underlying word has faded, such a word was treated as an underlying word. For this reason, *gyvas* was divided into *gy-v-as*, while *kuklus* into *kukl-us*.

Consonants tend to react actively to neighbouring sounds (assimilation, degemination, alternation, elimination of a consonant may occur); as a result, it is not always simple to recognise and to identify exact morpheme boundaries nor to name specific processes of consonant harmonisation (for more see Kazlauskienė and Cvilikaitė 2015). The future tense suffix *-s-* could be mentioned as an example: *bėg-s* 'he will run', *bėg-s-iu* 'I will run'. When a root ends on *s*, morphemes overlap and the remaining single consonant belongs to both morphemes, e.g., *ves+s=ves* 'he will marry', *vesiu* 'I will marry'.

This article employs the following terms: an **onset** is the initial consonant group of a morpheme, a **coda** is the final consonant group, and a **medial cluster** is a consonant group located between vowels in the middle of a morpheme. A **pattern** is considered to be a CV structure unit (with indicated quantity of consonants and a vowel or a diphthong) and a **type** is RTV (with an indicated consonant group according to the articulation thereof).

The Pattern Structure of Nominal Word Roots

The Structural Patterns of Roots

Lithuanian words and syllables can start and end on a vowel or a consonant (up to three consonants) (Girdenis & Karosienė, 2010, pp. 81–106; Kazlauskienė & Raškinis, 2008b, p. 26). There may be from 1 to 4 intervocalic consonants (Girdenis, 2014, pp. 130–132, Kazlauskienė & Raškinis, 2008a). A hiatus (a juncture of two vowels) in words of Lithuanian origin and old borrowings is only possible between morphemes; it is not possible between all morphemes, but only between a prefix and a root (e.g., *pa-akys* 'under eye') and in compounds, when a connecting vowel is added, e.g., *ilga-uodegė* 'long-tailed'. Based on structural regularities of other language units, the theoretic structure of a monosyllabic root could be as follows[3]: $C_{0-3}V(W)C_{0-3}$; the structure of a disyllabic root would be $C_{0-3}V(W)C_{1-4}V(W)C_{0-3}$, trisyllabic - $C_{0-3}V(W)C_{1-4}V(W)C_{1-4}V(W)C_{1-4}$ etc. However, such theoretic possibilities are not always materialized; furthermore, instances unforeseen in theoretic constructs also emerge.

The root structure of Lithuanian primary nominal words was extensively described by Karosienė (2004)[4]. The results of her research served as the basis to establish the inventory of structural patterns of nominal words. According to the research results of Karosienė, the roots of primary Lithuanian nominal words can be non-syllabic (there are only 10 roots), monosyllabic (78%), disyllabic (21%) and trisyllabic (15 roots). There may be no initial consonant group, while the final group, according to the research by Karosienė, is essential, and the total consonant amount in roots does not exceed 6. Thus the consonant number

3 The generalised formula should be read as follows: there may or may not be up to three consonants in an onset, the root centre can contain a vowel or a diphthong, there may or may not be up to three consonants in a coda.

4 The main source of Karosienė is the second edition of The Dictionary of Modern Lithuanian (futher in the text DML); some details were specified with reference to the third edition (Karosienė, 2004, p. 16).

provided in the formulas above is the maximum; however, it correlates with other root consonant clusters: if the onset has 3 consonants, the maximum number of consonants in the coda of monosyllabic roots can be 3, while in disyllabic roots respectively the sum of consonants of a medial cluster and a coda will not exceed 3. On the grounds of the research results of Karosienė, the theoretic formulas for root structure have to be revised considerably: $C_{0-3}V(W)C_{1-4}$ (Karosienė, 2004, p. 22), $C_{0-3}V(W)C_{1-4}V(W)C_{1-4}$, for trisyllabic roots $C_{0-1}VC_{1-2}VC_{1-2}VC_{1-2}$ (Karosienė, 2004, p. 73).

Based on the possible number of consonants, there can be 15 theoretic patterns of monosyllabic roots and 30 patterns of disyllabic roots. The research conducted by Karosienė revealed 46 realized patterns. All monosyllabic patterns are realized; while 8 patterns are not among the disyllabic realized ones (they are patterns with 5–6 consonants: VC_1VC_4, VC_2VC_{3-4}, VC_3VC_{2-3}, VC_4VC_2, $C_3VC_1VC_2$, $C_1VC_4VC_1$); only 5 patterns of trisyllabic roots are realized.

The structural patterns of roots of primary nouns and adjectives[5] are as follows:

$V(W)C_1$ – 41 (e.g., ***ak****is* ‘an eye’, ***ūm****us* ‘flash’);

$V(W)C_2$ – 35 (e.g., ***egl****ė* ‘a fir tree’, ***aišk****us* ‘clear’);

$V(W)C_3$ – 14 (e.g., ***ūksm****ė* ‘a shade’, ***aštr****us* ‘sharp’);

VC_4 – 6 (e.g., ***alksn****is* ‘an alder’, ***ankšt****as* ‘tight’);

$C_1V(W)C_1$ – 489 (e.g., ***bitė*** ‘a bee’, ***maž****as* ‘small’);

$C_1V(W)C_2$ – 320 (e.g., ***gerv****ė* ‘a crane’, ***gard****us* ‘tasty’);

$C_1V(W)C_3$ – 80 (e.g., ***vabzd****ys* ‘an insect’, ***karšt****as* ‘hot’);

C_1VC_4 – 14 (e.g., ***šerkšn****as* ‘frost’, ***linksm****as* ‘joyful’);

$C_2V(W)C_1$ – 192 (e.g., ***knyg****a* ‘a book’, ***stor****as* ‘fat’);

$C_2V(W)C_2$ – 105 (e.g. ***krašt****as* ‘an edge’, ***brang****us* ‘expensive’);

$C_2V(W)C_3$ – 30 (e.g., ***žvirbl****is* ‘a sparrow’, ***grakšt****us* ‘graceful’);

C_2VC_4 – 5 (e.g., ***slenkst****is* ‘threshold’);

C_3VC_1 – 12 (e.g., ***strėl****ė* ‘an arrow’, ***strop****us* ‘studious’);

$C_3V(W)C_2$ – 14 (e.g., ***skruzd****ė* ‘an ant’);

C_3WC_3 – 1 (e.g., ***straipsn****is* ‘an article’);

$V(W)C_1VC_1$ – 36 (e.g., ***erel****is* ‘an eagle’, ***atid****us* ‘attentive’);

VC_1VC_2 – 2 (e.g., ***agurk****as* ‘a cucumber’);

VC_1VC_3 – 1 (e.g., ***ielakšt****is* ‘a part of a plough’);

$V(W)C_2VC_1$ – 18 (e.g., ***elget****a* ‘a beggar’);

VC_2VC_2 – 1 (e.g., ***agrast****as* ‘a gooseberry’);

VC_3VC_1 – 3 (e.g., ***akstin****as* ‘an urge’);

VC_4VC_1 – 1 (e.g., ***inkštir****as* ‘a blackhead’);

$C_1V(W)C_1V(W)C_1$ – 179 (e.g., ***dešin****ė* ‘right’, ***malon****us* ‘pleasant’);

$C_1VC_1VC_2$ – 22 (e.g., ***baland****is* ‘a dove’);

$C_1VC_1V(W)C_3$ – 2 (e.g., ***žeberkl****as* ‘a spear’);

5 Based on the data of roots provided in the study by Karosienė (2004), this list contains root quantity and examples of a particular structural pattern. Lithuanian linguists tend to provide both, masculine and feminine forms, e.g., *geras, -a* ‘good’, for headwords of gender inflective words (e.g., adjectives, some pronouns). In this paper, for the sake of simplicity only the masculine gender will be provided.

$C_1VC_1VC_4$ – 4 (e.g., ***žebenkšt****is* ‘a weasel’);

$C_1VC_2VC_1$ – 49 (e.g., ***milžin****as* ‘a giant’, ***mandag****us* ‘polite’);

$C_1VC_2VC_2$ – 3 (e.g., ***serbent****as* ‘a current’);

$C_1VC_2VC_3$ – 1 (e.g., ***šermukšn****is* ‘a sorb’);

$C_1VC_3VC_1$ – 3 (e.g., ***garstyč****ia* ‘mustard’);

$C_2V(W)C_1VC_1$ – 22 (e.g., ***stubur****as* ‘a spine’);

$C_2VC_1VC_2$ – 4 (e.g., ***skiland****is* ‘Lithuanian meat specialty’, ***prabang****us* ‘luxurious’);

$C_2VC_1VC_3$ – 2 (e.g., ***skeveldr****a* ‘a shatter’, ***stamantr****us* ‘rigid’);

$C_2V(W)C_2VC_1$ – 10 (e.g., ***smilkin****ys* ‘a temple’);

$C_2VC_2VC_2$ – 1 (e.g., ***prašmatn****us* ‘fancy’);

$C_2VC_3VC_1$ – 1 (e.g., ***blakstien****a* ‘a lash’);

$C_3VC_1VC_1$ – 2 (e.g., ***skrybėl****ė* ‘a hat’);

$C_3VC_2VC_1$ – 1 (e.g., ***strazdan****a* ‘a freckle’);

$VC_1VC_2VC_1$ – 1 (e.g., ***atostog****os* ‘a holiday’);

$C_1VC_1VC_1VC_1$ – 10 (e.g., ***baravyk****as* ‘a boletus’);

$C_1VC_1VC_1VC_2$ – 1 (e.g., ***peteliškė*** ‘a butterfly’);

$C_1VC_1VC_2VC_1$ – 1 (e.g., ***dedervinė*** ‘herpes’);

$C_1VC_2VC_1VC_1$ – 2 (e.g., ***kankorėž****is* ‘a cone’).

Patterns $C_1V(W)C_1$, $C_1V(W)C_2$ and $C_2V(W)C_1$ are the most productive, as they make up more than a half of all roots of primary nominal words (55%). Thus, we can expect that these particular patterns will dominate in real connected speech.

What determines productivity of a pattern? First, attention should be paid to the quantity of consonants in a pattern: 3% of primary nominal words have 1 consonant, 36% contain 2 consonants, 37% have 3 consonants, 18% have 4 consonants, 5% have 5 consonants, and 1% has 6 consonants. It is obvious that the prevailing patterns are those whose total amount of consonants is between 2 and 3. Such are three most productive patterns mentioned above. However, the total amount of consonants is not the only factor influencing the productivity of a pattern. For example, the productivity of patterns $C_1V(W)C_2$ and $V(W)C_3$, both of which have three consonants, differs: respectively 14% and 1% of primary nominal words; the productivity of patterns $C_2V(W)C_2$ and $C_1V(W)C_3$ with four consonants is 7% and 4%. A more thorough analysis of pattern productivity suggests a conclusion that the productivity of a pattern largely depends on the quantity of consonants in the coda: the larger the number, the rarer the pattern. Patterns with 3–4 consonants in the coda do not tend to have high productivity; they represent respectively 8% and 1% of all primary nominal words. The dominating patterns are those with 1 consonant (they make up 64%) and 2 consonants (26%) in the coda. All these factors suggest that a complicated coda is not a usual phenomenon in the Lithuanian language and it would be the result of a juncture of a historical root and a consonantal suffix.

The variety of patterns for numerals and pronouns is not rich. They can be: C_1 (8 roots, e.g., ***d****u* ‘two’), C_2 (***tr****ys* ‘three’), $V(W)C_1$ (3 roots, e.g., ***aš*** ‘I’), VC_3 (***antr****as* ‘second’), $C_1V(W)C_1$ (13 roots, e.g., ***kit****as* ‘other’), $C_1V(W)C_2$ (3 roots, e.g., ***pirm****as* ‘first’), C_1VC_3 (***tamst****a* ‘Sir’), $C_1V(W)C_1V(W)C_1$ (2 roots, e.g., ***ketur****i* ‘four’), $C_1VC_1VC_2$ (***dešimt****is* ‘ten’), $C_1VC_2VC_1$ (2 roots, e.g., ***septyn****i* ‘seven’), $C_1VC_3VC_2$ (***tūkstant****is* ‘thousand’).

The analysis of primary roots is a good starting point to identify structural possibilities of morphemes. However, in real usage not only are particular primary words selected from the available inventory, but, also, derivatives are produced from primary words. Usage frequency of the latter may have influence on the frequency of some root patterns.

In DbML, we found 43 structural root patterns of nominal words. The major part of these patterns is within the periphery of usage; there are 3 main patterns: $C_1V(W)C_1$, $C_1V(W)C_2$, $C_2V(W)C_1$. Not only are they the most productive (the highest number of different roots in DML), but, also, their examples are most frequently used in real language (see Table 1). The dominating pattern in DbML is $C_1V(W)C_1$ nouns and adjectives which make up almost a half of all used examples. Approximately a quarter of all used examples are noun and adjective $C_1V(W)C_2$ and $C_2V(W)C_1$ structure roots. Thus the aforementioned structural patterns comprise even ¾ of all noun and adjective examples in DbML. Four more patterns, $C_1V(W)$, $V(W)C_1$, $C_2V(W)C_2$, $V(W)C_2$ and $C_1V(W)C_1V(W)C_1$, constitute almost a tenth of all examples. Noun and adjective roots with other patterns are rather rare in real language.

The numeral and pronoun group is dominated by non-syllabic C_1 roots (they are mainly pronouns, because there is only 0.2% of numerals with C_1 structure) and $C_1V(W)C_1$. These two patterns amount to 85% of all examples with numerals and pronouns in DbML.

Real usage highlighted such root patterns which were impossible to find during the analysis of lemmas (i.e., Sg Nom) and primary nominal words only. First, patterns whose roots end on a vowel (without a coda) should be mentioned; the majority of such nominal words were made from verbs, e.g., $C_1V(W)$ – ***bū****klė,* ***bū****sena* 'a state, condition' (cf. ***bū****ti* 'to be'), W – ***ei****smas* 'traffic' (cf. ***ei****ti* 'to go'), $C_3V(W)$ – *ap****skri****tas* 'round' (cf. *ap****skrie****ti* 'to orbit'). The other group encompasses scarce nominal words, e.g., VC_2VC_3 – ***atvirkšč****ias* 'reverse', $VC_1V(W)C_1VC_1$ – ***abėcėlė*** 'alphabet', or word forms, e.g., $C_1V(W)$ - ***sau*** (Dat 'oneself').

Table 1
The distribution of structural root patterns of nominal words in DbML[6]

Structural patterns	Usage instances		
	Nouns	Adjectives	Numerals and pronouns
C_1	0.1	0.5	53.4
C_2	0.02	0.7	2.6
$C_1V(W)C_1$	49.4	49.3	33.2
$C_1V(W)C_2$	18.6	11.4	2.4
$C_2V(W)C_1$	12.9	11.9	-
$C_1V(W)$*	3.4	3.1	0.3
$V(W)C_1$	3.0	2.7	4.3
$C_2V(W)C_2$	2.1	7.3	-
$V(W)C_2$	1.5	3.7	-
$C_1V(W)C_3$	1.4	2.5	0.1
$C_2V(W)$*	0.7	0.6	
W*	0.6	0.2	-
C_3VC_2	0.4	0.02	-
$C_3V(W)C_1$	0.4	0.1	-
$C_2V(W)C_3$	0.2	0.2	-

6 Symbol * is used to mark the patterns which were not mentioned by Karosienė; they can be identified only during the analysis of connected speech.

Structural patterns	Usage instances		
	Nouns	Adjectives	Numerals and pronouns
$V(W)C_3$	0.1	0.3	0.8
$C_3V(W)$*	0.1	0.1	-
VC_4	0.1	0.5	-
C_2VC_4	0.1	-	-
C_1VC_4	0.01	0.2	-
$C_1V(W)C_1V(W)C_1$	2.8	2.4	0.8
$C_1V(W)C_2VC_1$	0.5	0.8	0.5
$V(W)C_2VC_1$	0.4	0.1	0.3
$V(W)C_1VC_1$	0.4	-	-
$C_1VC_1VC_2$	0.3	1.1	1.1
$C_2V(W)C_1VC_1$	0.3	-	-
$C_2VC_1VC_2$	0.2	0.03	-
$C_1VC_1V(W)C_3$	0.1	-	-
$C_1VC_1VC_4$	0.02	-	
$C_1V(W)C_2VC_2$	0.02	-	-
$C_2VC_2VC_1$	0.01	-	-
VC_3VC_1	0.01	-	-
VC_1VC_2	0.01	-	-
VC_2VC_2	0.01	-	-
$C_1VC_2VC_3$	0.004	-	-
$C_1VC_3VC_1$	0.004	-	-
$C_3VC_1VC_1$	0.002	-	-
VC_2VC_3*	-	0.03	
$C_1VC_3VC_2$	-	-	0.4
$VC_1VC_2VC_1$	0.04	-	-
$C_1VC_1V(W)C_1VC_1$	0.04	-	-
$VC_1V(W)C_1VC_1$*	0.01	-	-
$C_1VC_1VC_1VC_2$	0.002	-	-

The beginning of a root may contain from 1 to 3 consonants. There are three types of binary clusters: ST (e.g., ***stor**as* 'fat'), SR (e.g., ***slenkst**is* 'threshold'), TR (e.g., ***knyg**a* 'a book'), with the latter type being the most common, as even half of consonant clusters of nominal word roots represent this type. The trinomial consonant cluster of a morpheme beginning is rare (only 4% of words in DbML) and it is only the STR pattern (e.g., ***skruzd**ė* 'an ant').

The variety of consonant clusters in the binary ending of a root and in a medial cluster is richer and embraces even 8 types: TT (e.g., ***daikt**as* 'a thing', ***dukter**s* 'daughter' Sg Gen), TS (e.g., ***lopš**ys* 'a cradle', ***aksom**as* 'velvet'), TR (e.g., ***kakl**as* 'a neck', ***septyn**i* 'seven'), ST (e.g., ***tašk**as* 'a dot', ***pastar**as* 'latter'), SR (e.g., ***dažn**as* 'frequent', ***paslik**as* 'prostrate'), RR (e.g., ***deln**as* 'a palm', ***karvel**is* 'a pigeon'), RT (e.g., ***smark**us* 'brisk', ***elget**a* 'a beggar'), RS (e.g., ***smals**us* 'curious', *mil**ž**inas* 'a giant'). However, the frequency of cluster types varies. The most common type is RT, amounting to a third of ending and medial clusters.

A trinomial ending of a root can be TST (e.g., ***šaukšt**as* 'a spoon'), TSR (e.g., ***sluoksn**is* 'a layer'), STR (e.g., ***meistr**as* 'a master'), RTT (e.g., ***virpt**is* 'hop'), RST (e.g., ***karšt**as* 'hot'), RSR (e.g., ***kremzl**ė* 'cartilage') and RTR (e.g., ***gandr**as* 'a stork'). A trinomial medial cluster is TST (e.g., ***tūkstant**is* 'thousand'), RST (e.g., ***erškėt**is* 'a thorn') and RTR (e.g., ***kerplėš**a* 'a snag'). The most frequent are TST, TSR and RTR (together they make up 8%) and TST medial clusters (10%).

Quaternary endings of root and medial clusters are rare: RTSR (e.g., **alksn**is 'an alder'), RTST (e.g., ***slenkst**is* 'threshold'), RSTR (e.g., ***iršt**va* 'a lie'), medial cluster RTST (the only example is ***inkštir**as* 'a blackhead', which might be a derivative).

The Structure of Verb Roots

The structure of roots of Lithuanian verbs provided in DML was described by Kruopienė (2000)[7]. The results of this analysis are used for the description of the inventory of structural patterns of primary verb roots.

According to the results of Kruopienė's research, the roots of Lithuanian primary verbs are only monosyllabic. This is their major difference from nominal words, which can be both non-syllabic and polysyllabic. However, like in roots of nominal words, there may be no onset in verbs, while the coda, based on Kuopienė's analysis, is compulsory[8], and the number of consonants in a root does not exceed 6, either.

In view of research results by Kruopienė, the structural formula for roots of primary verbs is as follows: $C_{0-3}V(W)C_{1-4}$ (Kruopienė, 2000, pp. 60–61, 100). There may be 15 theoretic patterns; however, only $C_3V(W)C_3$ pattern is not realised. The following patterns are the most productive: $C_1V(W)C_1$, $C_2V(W)C_1$, $C_1V(W)C_2$; together they make up 77% of all roots of primary verbs. For this reason, we can infer that the roots of such structure will dominate in real usage.

The structural patterns of primary verb roots:

$V(W)C_1$ – 20 (e.g., ***šė** 'he soughed');

$V(W)C_2$ – 14 (e.g., ***irz**o* 'he embittered');

$V(W)C_3$ – 5 (e.g., ***urzg**ė* 'he growled');

$V(W)C_4$ – 2 (e.g., ***inkšt**ė* 'he whined');

$C_1V(W)C_1$ – 392 (e.g., ***deg**ė* 'he burned');

$C_1V(W)C_2$ – 185 (e.g., ***valg**ė* 'he ate');

$C_1V(W)C_3$ – 15 (e.g., ***birzg**ė* 'he buzzed');

$C_1V(W)C_4$ – 4 (e.g., ***gergždė** 'he wheezed');

$C_2V(W)C_1$ – 258 (e.g., ***klaus**ė* 'he listened');

$C_2V(W)C_2$ – 120 (e.g., ***skalb**ė* 'he washed');

$C_2V(W)C_3$ – 26 (e.g., ***šniokšt**ė* 'he roared');

$C_2V(W)C_4$ – 19 (e.g., ***krenkšt**ė* 'he croaked');

$C_3V(W)C_1$ – 18 (e.g., ***skriej**o* 'he scudded');

$C_3V(W)C_2$ – 7 (e.g., ***spring**o* 'he choked').

The connection between pattern productivity and consonant number is similar to that of primary nominal words: 2% of primary verbs have 1 consonant, 37% have 2 consonants, 41%

7 The main data source of Kruopienė is the third edition of DML (Kruopienė, 2000, p. 5).

8 This was influenced by Kruopienė's choice not to analyse the roots of three lemmas of verbs (infinitive, present and past simple tenses) from the perspective of the modern language, but, rather, only historically non-derivative forms of present and past simple tenses, e.g., the root allomorphs of *eiti, eina, ėjo* 'to go, he goes, he went' are {ei-}, {ein-}, {ėj-}, the latter have codas, while the infinitive form does not.

have 3 consonants, 14% have 4 consonants, 3% have 5 consonants, and only 2% contain 6 consonants. Thus, patterns with 2 to 3 consonants dominate. Like in the group with nominal words, the patterns whose codas contain 1 consonant (64%) or 2 consonants (30%) are the most productive. There are 4% of trinary codas and only 2% of quaternary codas. Thus, a complex coda is certainly not a typical feature of the Lithuanian language.

All verb roots used in DbML were classified into 23 structural patterns. Based on the data from DbML, in coherent Lithuanian texts monosyllabic verb roots make up 99% and have $C_{0-3}V(W)C_{0-4}$ structure. The most frequent patterns $C_1V(W)C_1$, $C_2V(W)C_1$, $C_1V(W)C_2$ together amount to 78% of all verb roots in DbML (see Table 2).

Table 2
The distribution of structural patterns of verb roots in DbML (%)[9]

Structural patterns	Usage instances
$C_1V(W)C_1$	56.2
$C_1V(W)C_2$	11.9
$C_2V(W)C_1$	11.2
$V(W)C_1$	7.4
$C_1V(W)$*	6.1
C_2VC_2	2.7
$V(W)C_2$	0.9
$C_2V(W)$*	0.8
W*	0.6
C_1VC_3	0.5
C_3VC_1	0.3
C_3VC_2	0.2
C_2VC_3	0.1
C_2VC_4	0.04
C_3V*	0.04
VC_3*	0.01
$C_1VC_1VC_1$*	0.2
$C_2VC_1VC_1$*	0.02
VC_1VC_1*	0.01
VC_2VC_1*	0.01
$C_2VC_2VC_1$*	0.01
C_1*	0.01
C_2*	0.01

General regularities for dictionary and usage data are partially similar. The dominating pattern is $C_1V(W)C_1$. DML has 36% of such roots, while the usage data provide nearly one and a half times more roots. Pattern $C_2V(W)C_1$ in usage is twice as less frequent (11%) than in DML (23%). There are slightly fewer instances of pattern $C_1V(W)C_2$ in usage (17% in DML, almost 12% in DbML). The usage data have also highlighted the structural patterns of roots of derivative verbs. First, we should pay attention to the roots of infinitive and of derivative verbs having non-coda or non-syllabic patterns, e.g., W – ***ei****ti* 'to walk', C_1 in a verb *su**t**apatinti* 'to identify', which derived from a compound adjective *tapatus* 'identical', C_2 - a verb ***dv****ejintis* 'to double', which is made from a numeral. It is obvious that all verbs with disyllabic roots are derivatives: $C_1VC_1VC_1$ – ***vakar****ėjančio* 'the evening was coming' Ptcp Masc Sg Gen), $C_2VC_1VC_1$ – ***skeryč****iotis* 'to saw the air', VC_1VC_1 – ***ašar****oja* 'he cries', VC_2VC_1 – ***apval****inamas* 'being rounded' Pass Ptcp Masc Sg Nom, $C_2VC_2VC_1$ – *apsis**karmal**avo* 'he ragged himself'.

Consonant clusters in the middle of verb roots are rare (primary verbs, as it has already been mentioned, are monosyllabic and do not have medial clusters). Medial clusters in DbML occur only in a few derivatives: ***pasnink****auti* 'to fast', *pri**elget**auti* 'to beg', *apsi**skarmal**avę* 'ragged', ***apval****inamas* 'being rounded'.

Like in nominal words, verb roots match the structural patterns of a syllable: trinomial STR and binomial ST, SR, TR (e.g., ***spring****o* 'he choked', ***skalb****ė* 'he washed', ***šniokšt****ė* 'he roared', ***krenkšt****ė* 'he croaked'), with the latter cluster dominating (47% in DbML).

Endings of primary verbs are not as varied as in nominal words. Verbs tend to have only four binomial types of endings ST, RS, RT, RR, trinomial types TST, RST and quaternary RTST (e.g., respectively ***vystė*** 'he developed, ***dels****ė* 'he procrastinated', ***valg****ė* 'he ate', ***varv****a* 'it drops', ***šniokšt****ė* 'he roared', ***urzg****ė* 'he growled', ***gergžd****ė* 'he wheezed'). In real usage endings of verb

9 Symbol * is used to mark the patterns which were not mentioned by Kruopienė; they can be identified only during the analysis of connected speech.

roots are more diverse: without the aforementioned types, there were instances of SR (e.g., *pa**dažn**ėjo* 'it became frequent'), TR (e.g., ***liekn**ina* 'it slims'), TS (e.g., *pra**žiops**ojo* 'he missed something'), TT (e.g., ***žiopč**iodamas* 'gasping'), STR (e.g., *pa**aštr**ėjus* 'having exacerbated'), TSR (e.g., *pa**tiksl**inti* 'to specify'), RTT (e.g., ***bendr**auti* 'to communicate'), RTS (e.g., ***purkš**ti* 'to sprinkle'), RTR (e.g., *pa**lengv**inti* 'to alleviate'), RTSR (e.g., *ap**šerkšn**ijęs* 'frosted'). All these types are derivatives except the verb *purkšti*[10]. The dominating type of codas is RT (71% of real usage verbs); ST is also frequent (8% of real usage verbs).

Conclusions

The productivity and frequency analysis of morpheme structural patterns of roots of inflective parts of speech allows drawing the following conclusions.

1. 1,767 roots of primary nominal words can be divided into 46 structural patterns, and 1,085 roots of primary verbs into 14 patterns. This means that the relative structural diversity of nominal roots is higher than in verbs. This conclusion is confirmed by the relation between the numbers of different roots and patterns, that is 38 for nominal words and 78 for verbs[11].
2. Roots of primary nominal words are nonsyllabic C_{1-2}, monosyllabic $C_{0-3}V(W)C_{1-4}$, disyllabic $C_{0-3}V(W)C_{1-4}V(W)C_{1-4}$, trisyllabic $C_{0-1}VC_{1-2}VC_{1-2}VC_{1-2,}$ while roots of primary verbs are only monosyllabic: $C_{0-3}V(W)C_{1-4}$.
3. The total amount of root consonants does not exceed six. The most productive root patterns are $C_{1-2}V(W)C_{1-2}$ for nominal words and verbs (respectively 55% and 77%). The usage is dominated by root structures $C_{1-2}V(W)C_{1-2}$ for nouns, adjectives and verbs (77%) and numeral and pronoun roots C_1 and $C_1V(W)C_1$ (87%).
4. Binomial and trinomial clusters of all morpheme beginnings correspond to the patterns for syllable beginnings: STR, RT, SR, ST. Morphemes in medial clusters and codas may have all clusters of two consonants, except the cluster of two fricatives, which is impossible due to phonological processes inherent to the Lithuanian language. Trinomial consonant clusters in codas and medial clusters are very diverse (STR, TST, TSR, RTT, RST, RSR, RTR, RTS, RTSR, RTST, RSTR); however, consonant R never occurs in the middle of a cluster.

In conclusion, it should be said that the structural variety of morphemes is rich. However, the real usage is predominated by root patterns having a simple structure.

References

1. Akelaitienė, G., 1996. Morfonologinės balsių kaitos žodžių daryboje. [Morphonological vowel alternations in word formation]. Vilnius: Vilniaus pedagoginis universitetas.
2. Akelaitienė, G., 2000. Šaknies balsių kaita: morfonemų modeliai ir struktūriniai šaknies tipai. [Alternations of root vowels: morphoneme patterns and structural root types]. In: Žmogus ir žodis, 3(1), pp. 3–9.
3. DML – Dabartinės lietuvių kalbos žodynas [The Dictionary of Modern Lithuanian – DML], 2nd ed., (ed.) Kruopas, J., 1972. Vilnius: Mintis. (3rd ed., (ed.) S. Keinys, 1993. Vilnius: Mokslo ir enciklopedijų leidykla).
4. Girdenis, A., 2014. Theoretical Foundations of Lithuanian Phonology. Vilnius: Eugrimas.
5. Girdenis, A., & Karosienė, V., 2010. Bendrinės lietuvių kalbos statistinė struktūra:

10 This type did not occur among primary verbs only because the structure of infinitive roots was not analysed. In the Lithuanian language, metathesis of *sk, zg, šk, žg* occurs before consonants, e.g. *dreskia* 'he scratches' - *drėksti* 'to scratch', *mezga* 'he knits' - *megzti* 'to knit'.

11 Here the relation is calculated in the following way: the number of different morpheme examples is divided by the number of patterns, i.e., 1,767/46 = 38. This means that one pattern covers 38 different roots of nominal words.

fonologijos dalykai. [The statistical structure of Standard Lithuanian: phonological issues], Vilnius: Vilniaus universiteto leidykla.

6. Jakobson, R., 1962. Selected Writings, vol. 1. The Hague: Mouton.
7. Karosienė, V., 2004. Bendrinės lietuvių kalbos vardažodžio šaknies struktūra. [The root structure of nominal words in Standard Lithuanian]. Vilnius: Vilniaus universiteto leidykla.
8. Kaukienė, A., 1994. Lietuvių kalbos veiksmažodžio istorija. [The history of verbs of the Lithuanian language], part 1. Klaipėda: Klaipėdos universiteto leidykla.
9. Kaukienė, A., 2002. Lietuvių kalbos veiksmažodžio istorija. [The history of verbs of the Lithuanian language], part 2. Klaipėda: Klaipėdos universiteto leidykla.
10. Kazlauskienė, A. & Cvilikaitė, J., 2015. Review and Classification of Phonological and Morphonological Processes at the Juncture of Morphemes of Lithuanian. In: Contributions to Morphology and Syntax: Proceedings of the 4th Greifswald University Conference on Baltic Languages, (eds.) Kessler, S. & Judžentis, A., Berlin: Logos, pp. 9–26.
11. Kazlauskienė, A. & Raškinis, G., 2008a. Intervokaliniai priebalsiai: trinarės, keturnarės ir penkianarės grupės. [Intervocalic consonants: trinomial, quaternary and quinquenary groups]. In: Kalbų studijos, No. 13, pp. 51–59.
12. Kazlauskienė, A. & Raškinis, G. 2008b. Lietuvių kalbos fonologinio skiemens struktūrinių modelių dažnumas. [The frequency of structural patterns of phonological syllables in the Lithuanian language]. In: Žmogus ir žodis, 10(1), pp. 24–31.
13. Kazlauskienė, A., Raškinis, G., & Vaičiūnas, A., 2010. Automatinis lietuvių kalbos žodžių skiemenavimas, kirčiavimas, transkribavimas. [Automatic syllabification, accentuation and tanscription of Lithuanian words]. Kaunas: Vytauto Didžiojo universiteto leidykla.
14. Kruopienė, I., 2000. Bendrinės lietuvių kalbos veiksmažodžio šaknies struktūra. Daktaro disretacija. [Root structure of verbs in Standard Lithuanian. Doctoral dissertation]. Vilnius: Vilniaus universitetas.
15. Lietuvių kalbos morfemikos duomenų bazė. Retrieved from http://tekstynas.vdu.lt/page.xhtml?id=morfema-db
16. Rimkutė, E., Kazlauskienė, A., & Raškinis, G., 2011. Abėcėlinis lietuvių kalbos morfemikos žodynas. [The alphabetic morphemics dictionary of the Lithuanian language]. (Parts 1–3). Kaunas: Vytauto Didžiojo universiteto leidykla. Retrieved from http://vddb.laba.lt/obj/LT-eLABa-0001:B.03~2011~ISBN_978-9955-12-732-1 [Accessed 19 June 2017].
17. Trubetzkoy, N. S., 1931. Gedanken über Morphonologie. In: Travaux du Cercle Linguistique de Prague, No. 4, pp. 160–163.

Santrauka

Asta Kazlauskienė, Jurgita Cvilikaitė. Lietuvių kalbos kaitomųjų žodžių šaknies struktūra

Tyrimo tikslas – nustatyti lietuvių kalbos kaitomųjų žodžių šaknies struktūrinius modelius, jų produktyvumą ir dažnumą. Pirmiausia, remiantis ankstesniais lietuvių kalbininkų darbais, apžvelgiama struktūrinė šaknies morfemų įvairovė, nustatomas struktūrinių modelių produktyvumas (kiek yra konkretaus modelio skirtingų šaknų). Tada nagrinėjami ir aptariami realios vartosenos duomenys. Vartosenos duomenų tyrimui naudota *Lietuvių kalbos morfemikos duomenų bazė.* Tiriamąją medžiagą sudaro 265 tūkst. kaitomųjų žodžių pavartojimo atvejų.

Šaknies struktūros analizė leidžia daryti tokias išvadas: 1) morfemų struktūrinė įvairovė didelė, tačiau produktyvios ir dažnos yra nesudėtingos struktūros šaknys (produktyviausios ir dažniausios yra šaknys, kurių inicialėje ir finalėje yra nuo vieno iki dviejų priebalsių), 2) šaknies morfemos gali būti neskiemeninės ir nuo vieno iki trijų skiemenų (produktyviausios ir dažniausios yra vienskiemenės šaknys), 3) priebalsių samplaikos nėra dažnos šaknies

morfemų viduje (jos fiksuotos trečdalyje šaknų), 4) priebalsių kiekis šaknyje paprastai neviršija šešių, 5) priebalsių samplaikos inicialėje sutampa su skiemens pradžios modeliu, priebalsių samplaikos finalėje ir medialėje daug įvairesnės.

About the Authors

Asta Kazlauskienė

Prof. Dr., Professor at the Department of Lithuanian Studies at Vytautas Magnus University

Research interests

research of Lithuanian speech: the acoustic characteristics of sounds, phonotactics, accentuation and rythmics, applied linguistics

Address

V. Putvinskio g. 23-204, LT-44243 Kaunas, Lithuania

E-mail

asta.kazlauskiene@vdu.lt

Jurgita Cvilikaitė

Lecturer at the Institute of Foreign Languages at Vytautas Magnus University

Research interests

applied linguistics, foreign language teaching methodology, implementation of ICT in language teaching; corpus linguistics

Address

K. Donelaičio g. 52-603, LT-44243 Kaunas, Lithuania

E-mail

jurgita.cvilikaite@vdu.lt