

Productivity of Lithuanian Diminutives in Relation to Other Denominal and Deverbal Categories of Nouns¹

Jurgis Pakerys, Vilnius University, Lithuania

Virginijus Dadurkevičius, Vytautas Magnus University, Lithuania

Agnė Navickaitė-Klišauskienė, Vilnius University, Lithuania

Abstract

Lithuanian is one of the languages in which diminutives are claimed to be highly productive. In our study, we examine the productivity of three salient Lithuanian diminutive suffixes using a 1.3 billion-token corpus and propose that the degree of diminutive productivity in Lithuanian and other languages can be assessed by comparing it to the productivity of other denominal and deverbal categories of derived nouns. We conclude that the diminutive suffixes in Lithuanian are not as productive as those of action nominals but rank above quality, agent, and personal nouns.

Keywords: *diminutives, morphological productivity, corpus, Lithuanian.*

1 Introduction

Lithuanian belongs to the group of languages in which diminutives are highly productive and function as a salient category in language acquisition (Ulvydas 1965: 254; Ambrazas 1994: 87; Savickienė 2007; Savickienė et al. 2009; Dabašinskienė 2009; Dabašinskienė 2012; Dabašinskienė & Voeikova 2015; Stundžia 2016: 3095–3096; Dressler et al. 2022; Balčiūnienė et al. 2024: 394). As seen from these references, the role of Lithuanian diminutives has been mostly explored in language acquisition studies, while the data of adult language corpora are much less studied (Macienė 2005: 28; Inčiuraitė-Noreikienė 2017: 94).

In the present study, we assess the productivity of the most salient Lithuanian diminutive suffixes using a 1.3 billion-token corpus. We mostly use type and hapax counts, comparing them to data of other denominal and deverbal categories of derived nouns in order to determine the relative productivity of diminutives.

The paper is structured as follows. In Section 2, three productive diminutive suffixes of Lithuanian are introduced. In Section 3, the main corpus-based productivity measures are briefly reviewed. In Section 4, our corpus, lemmatizer, data extraction and review procedures are presented. In Section 5, the productivity measures of diminutives and other major deverbal and denominal categories of nouns are presented and discussed, and in Section 6, some conclusions are offered.

2 Lithuanian diminutive suffixes

We studied the following three diminutive suffixes that are listed as the most productive in the main grammars: *-el-*, *-ėl-*, and *-(i)uk-* (Ulvydas 1965: 254–269; Ambrazas 1994: 88–91; cf.

¹ This paper is based on the presentation given at the Word-Formation Theories VII & Typology and Universals in Word-Formation VI conference, held at Pavol Jozef Šafárik University in Košice, on June 28, 2025. The authors sincerely thank the audience of the conference and two anonymous reviewers for their constructive feedback and Cristina Aggazzotti for editing the English of the article.

Balčiūnienė et al. 2024: 394–395). The number of suffixes was limited to three due to the need for additional semi-automatic lemmatization and time-consuming manual review (see Section 4 for details). We acknowledge that including more suffixes of the diminutives and other derivational categories discussed in Section 5 may alter the results.

Lithuanian nouns have two genders, masculine and feminine: the diminutives with the nom. sg. endings *-as* and *-is* are masculine, while those with *-ė* are feminine, as seen in (1)–(8) below. The diminutive suffixes are mostly transparent to gender, as illustrated in (1)–(8), but some exceptions can occur, e.g., *up-ė* (f.) ‘river’ → *up-el-is* (m.) ‘stream, rivulet’.

The suffixes *-el-* and *-ėl-* are exceptional in that they are found in complementary distribution with respect to the length of the base in standard Lithuanian (Ulvydas 1965: 260, 264; Ambrazas 1994: 90; Balčiūnienė et al. 2024: 395); consider (1)–(2) with monosyllabic bases and (3)–(4) with polysyllabic ones. We will present the productivity measures for *-el-* and *-ėl-* separately for the sake of convenience.

- | | | | |
|-----|-----------------------------------|---|---|
| (1) | <i>vaik-as</i> (m.) ‘child’ | → | <i>vaik-el-is</i> (m.) ‘small child’ |
| (2) | <i>šak-a</i> (f.) ‘branch’ | → | <i>šak-el-ė</i> (f.) ‘small branch’ |
| (3) | <i>katin-as</i> (m.) ‘cat (male)’ | → | <i>katin-ėl-is</i> (m.) ‘small cat, kitten’ |
| (4) | <i>parduotuv-ė</i> (f.) ‘shop’ | → | <i>parduotuv-ėl-ė</i> (f.) ‘small shop’ |

The suffix *-(i)uk-* is insensitive to the length of the base, as seen in (5)–(8): the bases of (5)–(6) are monosyllabic, while those of (7)–(8) are disyllabic. This suffix is found in two variants: the one that induces non-automatic palatalization of the last consonant of the base, as in (6) (here, the non-palatalized final [l] <l> contrasts with the palatalized [lʲ] spelled as in the derivative)², and the one without the palatalization, as in (5), (7), and (8). For the sake of simplicity, we present productivity measures for the palatalizing and non-palatalizing variants of this suffix together.

- | | | | |
|-----|-----------------------------------|---|---|
| (5) | <i>rat-as</i> (m.) ‘wheel’ | → | <i>rat-uk-as</i> (m.) ‘small wheel’ |
| (6) | <i>suol-as</i> (m.) ‘bench’ | → | <i>suoli-uk-as</i> (m.) ‘small bench’ |
| (7) | <i>katin-as</i> (m.) ‘cat (male)’ | → | <i>katin-uk-as</i> (m.) ‘small cat, kitten’ |
| (8) | <i>kavin-ė</i> (f.) ‘cafe’ | → | <i>kavin-uk-ė</i> (f.) ‘small cafe’ |

Lithuanian diminutives are frequently derived from concrete nouns, both animate and inanimate, and their senses can typically be characterized by a reduced size in comparison to the standard denoted by the base, as illustrated above in (1)–(8); these formations are also used to express a positive attitude (fascination, endearment) towards the referent. Diminutives are also derivable from mass nouns, e.g., *vand-uo* (non-nominative stem *vanden-*) ‘water’ → *vanden-ėl-is*, *vanden(i)-uk-as*; *smėl-is* ‘sand’ → *smėli-uk-as*. Abstract nouns may also serve as bases, e.g., *klausim-as* ‘question’ → *klausim-ėl-is*, *klausim-uk-as*; *reikal-as* ‘matter, affair’ → *reikal-ėl-is*, *reikali-uk-as*. These diminutives, and also a number of the ones mentioned earlier, may carry a pragmatic function of mitigating the subject matter of the conversation; they may also express irony. Further details regarding the semantic and pragmatic aspects of Lithuanian diminutives can be found in Ulvydas (1965: 254–288), Ambrazas (1994: 87–94), Macienė

² If the last consonant of the base is [d] or [t], it is affricatized to [tʃ] and [dʒ], respectively, and spelled as <či> and <dži>; when segmenting the derivatives, we leave the letter <i>, indicating palatalization, as part of the base before the hyphen that separates the suffix, as in (6).

(2005), Savickienė (2007), Dabašinskienė (2009), Dabašinskienė & Voeikova (2015), and Dressler et al. (2022).

3 Corpus-based productivity measures

For estimating derivational productivity in corpora, the most widely used measures are realized, potential, and expanding productivity (Baayen 1992, 1993; see overviews in Baayen 2009, Gaeta & Ricca 2006: 844–849, and Dal & Namer 2016: 73–76). Realized productivity simply reflects the number of derived words with a given affix; this way of estimating derivational productivity was also used before the advent of corpus linguistics. The figures of realized productivity may be problematic, as this measure is unable to distinguish the processes that were productive in the past from the ones that are productive in the present. To solve this problem, measures involving hapaxes (lexemes attested only once in a given corpus) were introduced: expanding and potential productivity. The logic behind this is that in a comparatively large corpus, the pool of hapaxes contains a significant number of newly coined words. Expanding productivity either simply indicates the number of hapaxes with a certain affix or the relation of that number to the total number of hapaxes in a given corpus. Expanding productivity lets us determine the probability of finding a new formation with a certain affix in the pool of hapaxes. For the sake of simplicity, we will use the number of hapaxes found with our diminutive suffixes because determining a true number of hapaxes in our corpus is difficult due to the limited lemmatization capabilities discussed in Section 4.

In the case of potential productivity, the ratio of hapaxes with a given affix to the total number of tokens (= total frequency of grammatical forms) with a given affix is calculated. This determines the probability of finding a new formation as we sample through the pool of tokens derived with the affix of interest to us. The problem with potential productivity is that affixes with comparatively low total frequencies (in relation to other affixes under study) may be overestimated (van Marle 1992; Gaeta & Ricca 2006). A solution to this problem is either to use the variable-corpus approach (Gaeta & Ricca 2006) or to employ Large Number of Rare Events (LNRE) models (Evert & Baroni 2007; see application in, e.g., Štichauer 2009, Varvara 2019). At the present state of our study, we were unfortunately unable to use these methods and decided to resort to the original estimation and limited use of potential productivity. As seen in Section 5, we use the estimates of potential productivity only when diminutive suffixes are compared among each other; for cross-category comparison, we limit ourselves to only realized and expanding productivity.

4 Our corpus, lemmatization, and data review

We used the Joint Corpus of Lithuanian (JCL) containing 1.3 billion tokens and comprising the three subcorpora listed in Table 1. The lemma and word-form lists of the JCL are openly published (Dadurkevičius 2020a, 2020b), but the corpus as a whole is not available for immediate retrieval of information, which slows down the process of manual review. The main reason behind our choice of the JCL was the open access word-form lists that allowed us to perform additional semi-automatic lemmatization. We readily admit that the JCL is far from a well-balanced corpus, but we believe that the productivity measures obtained from that corpus are in general valid. This is partly confirmed by the productivity rankings of nominal suffixes

in the JCL when they correspond to linguistic intuitions reflected in the rankings of the main grammars, see Pakerys et al. (2024, 2025). However, this is not always the case: the grammars may be wrong in their estimation or our rankings may be affected by the specific composition of the JCL. Future studies should estimate the productivity of diminutives in corpora with better representation of spontaneous speech and informal registers where diminutives are expected to be more frequent.

Table 1: Subcorpora of the Joint Corpus of Lithuanian.

Subcorpus	Online access	Tokens
Lithuanian internet texts, Vilnius University, 2014	Not available	779,154,268
Legal documents, courtesy of the Office of the Seimas of the Republic of Lithuania, 2011	https://e-seimas.lrs.lt/portal/documentSearch/lt	443,114,936
Balanced Corpus of Modern Lithuanian, Vytautas Magnus University, 2008	http://tekstynas.vdu.lt/tekstynas/	112,575,876
Total		1,334,845,080

The lemmatizer that was used for the compilation of the JCL lemma list has a fixed dictionary and a set of inflectional rules (Dadurkevičius 2017). As a result, many diminutive formations that are not found in that dictionary were not recognized and not lemmatized. It is very well known that the quality of lemmatization and subsequent manual annotation significantly affects productivity measures (Evert & Lüdeling 2001; Gaeta & Ricca 2006; Dal et al. 2008; Baayen 2009). To capture the diminutives that were not recognized by our lemmatization tool, we performed the following semi-automatic lemmatization procedure. First, we filtered unannotated (non-lemmatized) corpus tokens according to the pattern DIMINUTIVE SUFFIX + (all possible) ENDINGS and grouped those tokens into potential lemmas. Then, we manually reviewed the resulting lists and removed derivationally non-transparent formations and non-existent lemmas that were automatically created in our first step due to homographic forms; we also removed lemmas that contained spelling errors. The additional semi-automatic lemmatization greatly improved the total lemma counts and hapax counts, as seen in Table 2. The manual review of potential lemmas was time-consuming and thus reduced our ability to include more suffixes in this study. Also, due to time limitations, we could not include diminutives derived from proper names; such an inclusion would have increased productivity values to a certain extent.

Table 2: Diminutive lemma counts after the two lemmatization stages (ranked according to the sum of masculine and feminine types).

Suffix	Automatic lemmatization		Automatic and additional semi-automatic lemmatization	
	Types	Hapaxes	Types	Hapaxes
-(i)uk-as m.	1,182	7	3,598	950
-(i)uk-ė f.	147	0	686	209
-ėl-is m.	483	3	1,975	577
-ėl-ė f.	302	2	1,064	264
-el-is m.	481	2	755	80

5 Productivity of diminutives with respect to other categories

First, let us analyze the productivity data of diminutives in Table 3, where token frequency and potential productivity values are presented, in addition to the data found in the last two columns of Table 2³.

Table 3: Productivity measures of diminutive suffixes (ranked according to the sum of masculine and feminine types).

Suffix	Types	Hapaxes	Tokens	Potential productivity ($\times 10^3$)
-(<i>i</i>) <i>uk-as</i> m.	3,598	950	687,339	1.38
-(<i>i</i>) <i>uk-ė</i> f.	686	209	24,372	8.58
- <i>ėl-is</i> m.	1,975	577	376,210	1.53
- <i>ėl-ė</i> f.	1,064	264	203,848	1.30
- <i>el-is</i> m.	755	80	1,354,266	0.06
- <i>el-ė</i> f.	651	72	1,098,361	0.07

The suffix *-(i)uk-* tops the ranking according to the type counts, followed by *-ėl-* and then *-el-*. This is in contrast to what the main grammars suggest, where *-el-* is listed first, followed by *-ėl-* and then *-(i)uk-* (Ulvydas 1965: 254–265; Ambrazas 1994: 88–90; Balčiūnienė et al. 2024: 394). A study of diminutives based on one of the corpora that is included in the JCL, namely the Balanced Corpus of Modern Lithuanian (BCML), reports the following data (Macienė 2005: 28): *-ėl-* and *-(i)uk-* have similar values (1,101 and 1,052 types, respectively), while *-el-* ranks lower (863 types), just as in our study. Another study limited to only the derivation of nouns with borrowed neoclassical roots in the BCML notably arrived at the same ranking as ours: *-(i)uk-* > *-ėl-* > *-el-* (Inčiuraitė-Noreikienė 2017: 94). All in all, we conclude that the grammars appear to overestimate the realized productivity of *-el-* and underestimate that of *-(i)uk-*. As one of the anonymous reviewers suggests, the claims of the grammars may rest on the token frequencies: this is in line with the JCL token frequency of *-el-*, which ranks first, but the suffix *-ėl-* still ranks lower than *-(i)uk-*; the differences may also stem from the texts used for the ranking of the grammars (fiction) compared to those included in our corpus.

According to expanding productivity, the ranking is the same. It is notable that the masculine and feminine formations in *-el-* have rather similar numbers of hapaxes, while *-ėl-* shows a more noticeable gap, and *-(i)uk-* has the biggest difference. In the case of nouns that refer to animates with gender distinctions, this creates the potential of forming feminine diminutives once communicative need arises. At the present stage of our research, we were unable to determine the proportion of diminutives that have animate referents with relevant gender distinctions in our data. We should note, however, that for the suffix *-(i)uk-*, there is simply a strong tendency to use masculine bases, as noted in the grammars (Ulvydas 1965: 268; Ambrazas 1994: 91). The reason behind this is unclear, but this tendency may explain the

³ Some figures differ slightly from the ones presented in Pakerys et al. (2025: 87) due to the updated version of the dataset.

significantly lower proportion of feminine types among *-(i)uk-* formations; the tendency may also be diminishing the potential to form new derivatives reflected in the number of hapaxes.

The values of potential productivity need to be interpreted with caution: the token frequencies range from 24,372 to 1,354,266, and, as mentioned in Section 3, the productivity of suffixes with comparatively low frequencies may be overestimated. A case in point appears to be the feminine suffix *-(i)uk-ė*: we have observed that there is a potential to form new feminine diminutives for animates with gender distinctions, but the potential productivity value of *-(i)uk-ė* still appears to be overestimated. The measure of potential productivity is also blind to certain general restrictions, as in the case of *-(i)uk-*, where the tendency to form masculine diminutives is observed. The low values of potential productivity for *-el-* formations, however, appear to be indicative: the pool of their bases is limited (recall that they need to be monosyllabic), while their total frequencies are high, apparently due to a comparatively large number of established (lexicalized) formations⁴. As for the potential productivity of the other suffixes (masculine ones in *-(i)uk-as* and masculine and feminine ones in *-ėl-*), we believe that they need to be reassessed by employing the variable-corpus approach or LNRE models mentioned in Section 3.

Now let us look at how the productivity of diminutives compares to that of other categories of nouns. In Table 4, we have selected two to three of the most productive suffixes in each category based on the judgment of the major grammars (and the results obtained in our earlier studies) for action nominals, quality nouns, agent nouns, and personal nouns. Less productive categories, such as status nouns, instrument nouns, result and object nouns, place nouns, etc. were excluded from the comparison. As mentioned earlier (Section 1), we acknowledge that the inclusion of more suffixes of diminutives and other categories may alter the results, but we leave this problem for future studies.

For the sake of clarity, we illustrate each suffix with one example below: deverbal action nominals in (9), deadjectival⁵ quality nouns in (10), deverbal agent nouns in (11), and denominal and deadjectival personal nouns in (12). Differently from the descriptive tradition of Lithuanian grammars, we only included the data for deadjectival quality nouns illustrated in (10) and separated them from other abstracts, such as denominal status nouns (e.g., *vaik-as* ‘child’ → *vaik-yst-ė* ‘childhood’), as done by Stundžia (2016: 3095–3096). The traditional approach, however, is to present both categories together under the broad label of *nomina qualitatis*. As seen from (9) and (12), we also include the suffixes *-acij-a* and *-ist-as* found in borrowings such as *registracija* ‘registration’ and *gitaristas* ‘guitarist’ that historically were not formed in Lithuanian. They can be interpreted as derived due to their synchronic relation to other borrowings that serve as their bases; consider *registr-uo-ti* ‘register’ and *gitar-a* ‘guitar’, which allow us to segment the nouns *registr-acij-a* and *gitar-ist-as* as having the suffixes *-acij-a* and *-ist-as*. It should be noted that these suffixes are only rarely used with native bases to produce new formations in Lithuanian.

- | | | | |
|-----|---------------------------------|---|--------------------------------------|
| (9) | <i>bėg-ti</i> ‘run’ | → | <i>bėg-im-as</i> ‘running’ |
| | <i>skait-y-ti</i> ‘read’ | → | <i>skait-ym-as</i> ‘reading’ |
| | <i>registr-uo-ti</i> ‘register’ | → | <i>registr-acij-a</i> ‘registration’ |

⁴ One of the anonymous reviewers offers a valuable observation that due to the monosyllabic bases, formations in *-el-* are shorter than the other diminutives and may become more frequent due to Zipf’s principle of least effort.

⁵ “Deadjectival” here and elsewhere refers to bases that are adjectives and adjective-like parts of speech (participles, ordinal numerals, some pronouns).

- (10) *sald-us* (m.), *-i* (f.) ‘sweet’ → *sald-um-as* ‘sweetness’
gaus-us (m.), *-i* (f.) ‘abundant’ → *gaus-yb-ė* ‘abundance’
- (11) *raš-y-ti* ‘write’ → *rašy-toj-as* (m.), *-a* (f.) ‘writer’
šok-ti ‘dance’ → *šok-ėj-as* (m.), *-a* (f.) ‘dancer’
bėg-ti ‘run’ → *bėg-ik-as* (m.), *-ė* (f.) ‘runner’
- (12) *men-as* ‘art’ → *men-inink-as* (m.), *-ė* (f.) ‘artist’
gitar-a ‘guitar’ → *gitar-ist-as* (m.), *-ė* (f.) ‘guitar player’
jaun-as (m.), *-a* (f.) ‘young’ → *jaun-uol-is* (m.), *-ė* (f.) ‘young man, woman’

Table 4 ranks the selected derivational categories according to the sums of the types of the most productive suffixes from each category alongside hapax and token sums. The token numbers are provided only to give an estimate of how frequently each category is used in the running text, while the potential productivity values would be worth considering only for the individual suffixes and are not presented here.

Table 4: Productivity of diminutives and other deverbal and denominal categories of nouns (ranked according to types).

Category	Suffixes	Types (totals)	Hapaxes (totals)	Tokens (totals)
Action nominals	<i>-im-as</i> <i>-ym-as</i> <i>-acij-a</i>	18,180	3,275	58,561,730
Diminutives	<i>-(i)uk-as, -ė</i> <i>-ėl-is, -ė</i> <i>-el-is, -ė</i>	8,610	2,121	3,744,398
Quality nouns	<i>-um-as</i> <i>-yb-ė</i>	6,793	1,577	8,900,847
Agent nouns	<i>-toj-as, -a</i> <i>-ėj-as, -ė</i> <i>-ik-as, -ė</i>	5,796	725	18,329,294
Personal nouns	<i>-inink-as, -ė</i> <i>-ist-as, -ė</i> <i>-uol-is, -ė</i>	5,033	1,342	6,054,612

The category of diminutives ranks second after the action nominals, which is in line with the general claims and linguistic intuitions that diminutives are highly productive in Lithuanian. It also seems to reflect the interpretation of diminutives as a category that is only partly lexicon-enriching and has frequent pragmatic and discourse-oriented uses. However, we consider this explanation only as an informed hypothesis; to what extent such pragmatic and discourse-related aspects affect the productivity of diminutives in Lithuanian should be explored in a separate study. The same can be also said of other categories included in our study: they significantly differ in their general semantics and use in discourse organization, but a more detailed characterization and quantification of each category was beyond the scope of the present study.

Deadjectival quality nouns are found in third place, which is related to our decision to separate them from the status nouns and other related formations with abstract semantics. Counting denominal status nouns with the suffix *-yst-ė* and formations with the suffix *-izm-as* (mostly borrowings) found in status and other abstract nouns⁶ together with deadjectival quality nouns as a joint category of abstract nouns would add 1,866 types and 566 hapaxes. The totals of 8,659 types and 2,143 hapaxes of this category would guarantee a shared second place with diminutives.

The next natural step of the analysis would be to compare the productivity data of Lithuanian diminutives with that of other languages. Unfortunately, the closest set of data we were able to find was only from one language, namely Dutch, provided in Baayen (1993). Due to the large difference in corpus size (1.3 billion tokens of Lithuanian vs. 600,000 tokens of Dutch), the results are hardly comparable, and the notes below should be treated as exploratory and provisional. In the said study, Dutch diminutives in *-tje* surpass the action nominals in *-ing* and quality nouns in *-heid* with respect to type and hapax counts; *-tje* also has a higher potential productivity value than the latter suffixes. Of the other categories included in that study, only compounds surpass diminutives in *-tje* by having larger numbers of types and hapaxes but a slightly lower potential productivity value. For the sake of comparison, we decided to select the most productive suffixes from our set of derivational categories and present them in Table 5.

Table 5: The most productive suffixes of action nominals, quality nouns, and diminutives (ranked according to types).

Category	Suffix	Types	Hapaxes	Tokens	Potential productivity ($\times 10^3$)
Action nominal	<i>-im-as</i>	15,274	2,855	41,351,162	0.07
Quality noun	<i>-um-as</i>	6,043	1,612	4,027,571	0.40
Diminutive	<i>-(i)uk-as</i> m.	3,598	950	687,339	1.38
	<i>-(i)uk-ė</i> f.	686	209	24,372	8.58

In Table 5, the ranking of categories changes: the most productive diminutive suffix *-(i)uk-* is surpassed by the quality noun suffix *-um-as*. This is interesting and needs to be explored further by comparing Dutch and Lithuanian corpora of similar sizes to determine how the productivity rank of Dutch diminutives changes as the corpus size increases. Finally, the potential productivity value is also of interest: the Lithuanian diminutive suffix *-(i)uk-* has the highest values, but due to vast differences among the frequencies of the suffixes under comparison, this result needs to be addressed with caution and reexamined by applying the variable-corpus approach or LNRE models.

6 Conclusions

We have proposed to evaluate the productivity of diminutives by comparing them to other salient derivational categories of nouns. Our study shows that Lithuanian diminutive suffixes

⁶ For example, *idiot-as* ‘idiot’ $\rightarrow$ *idiot-izm-as* ‘idiocy’, *kub-as* ‘cube’ $\rightarrow$ *kub-izm-as* ‘cubism’, *Marksas* ([Karl] Marx) $\rightarrow$ *marks-izm-as* ‘Marxism’, etc.

are not as productive as those of action nominals, but rank above quality, agent, and personal nouns. The high productivity of diminutives is probably related to their pragmatic and discourse-oriented uses that need to be explored in future studies. We also admit that our results may depend on the number of suffixes that are included for comparison; the inclusion of diminutives of proper names is also desirable. Different corpora, especially ones with a better representation of spontaneous spoken discourse and informal registers, may also yield different results. With regard to estimating potential productivity, the data have to be reexamined by applying the variable-corpus approach or LNRE models.

For cross-linguistic comparison, more data from similar corpora, across diverse languages, and with the same productivity measures are needed. The potential use of parallel translation corpora should also be explored. As for corpus annotation, we note that close attention needs to be paid to the quality of lemmatization and, when necessary, additional steps must be taken to achieve maximal numbers of lemmatized units.

Abbreviations

f – feminine
gen – genitive
m – masculine
sg – singular

References

- Ambrasas, Vytautas (ed.). 1994. *Dabartinės lietuvių kalbos gramatika*. Vilnius: Mokslo ir enciklopedijų leidykla.
- Baayen, R. Harald. 1992. Quantitative aspects of morphological productivity. In Booij, Geert E. & van Marle, Jaap (eds.), *Yearbook of Morphology 1991*, 109–149. Dordrecht: Kluwer. (https://doi.org/10.1007/978-94-011-2516-1_8)
- Baayen, R. Harald. 1993. On frequency, transparency and productivity. In Booij, Geert E. & van Marle, Jaap (eds.), *Yearbook of Morphology 1992*, 181–208. Dordrecht: Kluwer. (doi:10.1007/978-94-017-3710-4_7)
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Lüdeling, Anke & Kytö, Merja (eds.), *Corpus Linguistics: An International Handbook*, vol. 2, 899–919. Berlin, New York: De Gruyter Mouton. (doi:10.1515/9783110213881.2.899)
- Balčiūnienė, Asta & Drukeinis, Albinas & Kazlauskaitė, Rūta & Vaskelienė, Jolanta. 2024. *Praktinė bendrinės lietuvių kalbos gramatika*. Vilnius: Vilniaus universiteto leidykla.
- Dabašinskienė, Ineta. 2009. Intimacy, familiarity and formality: Diminutives in Modern Lithuanian. *Lituanus* 55(1). 65–80.
- Dabašinskienė, Ineta. 2012. Gender differences in language acquisition: A case study of Lithuanian diminutives. *Journal of Baltic Studies* 43(2). 177–196. (doi:10.1080/01629778.2012.674795)
- Dabašinskienė, Ineta & Voeikova, Maria. 2015. Diminutives in spoken Lithuanian and Russian: Pragmatic functions and structural properties. In Arkadiev, Peter & Holvoet, Axel & Wiemer, Björn (eds.), *Contemporary Approaches to Baltic Linguistics*, 203–234. Berlin, München, Boston: De Gruyter. (doi:10.1515/9783110343953-005)

- Dadurkevičius, Virginijus. 2017. Lietuvių kalbos morfologija atvirojo kodo „Hunspell“ platformoje [Lithuanian morphology in the Hunspell framework]. *Bendrinė kalba* 90. [1–17] (<https://journals.lki.lt/bendrinekalba/article/view/156>).
- Dadurkevičius, Virginijus. 2020a. Wordlist of Lemmas from the Joint Corpus of Lithuanian. CLARIN-LT digital library in the Republic of Lithuania. (<http://hdl.handle.net/20.500.11821/41>)
- Dadurkevičius, Virginijus. 2020b. Assessment Data of the Dictionary of Modern Lithuanian versus Joint Corpora. CLARIN-LT digital library in the Republic of Lithuania. (<https://clarin.vdu.lt/xmlui/handle/20.500.11821/36>)
- Dal, Georgette & Fradin, Bernard & Grabar, Natalia & Namer, Fiammetta & Lignon, Stéphanie & Plancq, Clément & Zweigenbaum, Pierre & Yvon, François. 2008. Quelques pré-alables au calcul de la productivité des règles constructionnelles et premiers résultats. In Durand, Jacques & Habert, Benoît & Laks, Bernard (eds.), *Actes du premier Congrès mondial de linguistique française, Paris, 9–12 juillet 2008*, 1587–1599. Paris: Institut de Linguistique Française. (doi:10.1051/cmlf08184)
- Dal, Georgette & Namer, Fiammetta. 2016. Productivity. In Hippius, Andrew & Stump, Gregory (eds.), *The Cambridge Handbook of Morphology*, 70–90. Cambridge: Cambridge University Press. (doi:10.1017/9781139814720.004)
- Dressler, Wolfgang U. & Mattiello, Elisa & Korecky-Kröll, Katharina & Noccetti, Sabrina & Dabašinskienė, Ineta & Kamandulytė-Merfeldienė, Laura & Kazakovskaya, Victoria V. 2022. Communication with diminutives to young children vs. pets in German, Italian, Lithuanian, Russian, and English. *Stem-, Spraak- en Taalpathologie* 27. (doi:10.21827/32.8310/2022-SG-53)
- Evert, Stefan & Baroni, Marco. 2007. zipfR: Word frequency distributions in R. In Ananiadou, Sophia (ed.), *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Demo and Poster Sessions*, 29–32. Prague, Czech Republic: Association for Computational Linguistics. (<https://aclanthology.org/P07-2008>)
- Evert, Stephanie & Lüdeling, Anke. 2001. Measuring morphological productivity: Is automatic preprocessing sufficient? In Rayson, Paul & Wilson, Andrew & McEnery, Tony & Hardie, Andrew & Khoja, Shereen (eds.), *Proceedings of the Corpus Linguistics 2001 Conference*, 167–175. Lancaster: Lancaster University.
- Gaeta, Livio & Ricca, Davide. 2006. Productivity in Italian word formation: a variable-corpus approach. *Linguistics* 44(1). 57–89. (doi:10.1515/LING.2006.003)
- Inčiuraitė-Noreikienė, Lina. 2017. *Dabartinės lietuvių kalbos vardažodžių su (neo)klasikiniais elementais morfeminė sandara ir daryba*. Doktoro disertacija. Vilnius: Vilniaus universiteto leidykla.
- Macienė, Jurgita. 2005. *Deminutyvų stilistika*. Šiauliai: Šiaulių universiteto leidykla.
- Pakerys, Jurgis & Navickaitė-Klišauskienė, Agnė & Dadurkevičius, Virginijus. 2024. Productivity of deverbal suffixal nouns in the Joint Corpus of Lithuanian. *Baltistica* 59(1). 5–39. (doi:10.15388/baltistica.60.1.2567)
- Pakerys, Jurgis, & Navickaitė-Klišauskienė, Agnė & Dadurkevičius, Virginijus. 2025. Productivity of denominal and deadjectival suffixal nouns in the Joint Corpus of Lithuanian. *Baltistica* 60(1). 75–94. (doi:10.15388/baltistica.59.1.2542)
- Savickienė, Ineta & Dressler, Wolfgang U. (eds.). 2007. *The Acquisition of Diminutives: A cross-linguistic perspective*. Amsterdam: John Benjamins. (doi:10.1075/lald.43)

- Savickienė, Ineta. 2007. Form and meaning of diminutives in Lithuanian child language. In Ineta Savickienė & Wolfgang U. Dressler (eds.), *The Acquisition of Diminutives: A cross-linguistic Perspective*, 13–41. Amsterdam: John Benjamins. (doi 10.1075/lald.43.02sav)
- Savickienė, Ineta & Kempe, Vera & Brooks, Patricia J. 2009. Acquisition of gender agreement in Lithuanian: Exploring the effect of diminutive usage in an elicited production task. *Journal of Child Language* 36(3). 477–494. (doi:10.1017/S0305000908009100)
- Štichauer, Pavel. 2009. Morphological productivity in diachrony: The case of the deverbal nouns in *-mento*, *-zione* and *-gione* in Old Italian from the 13th to the 16th century. In Montermini, Fabio & Boyé, Gilles & Tseng, Jesse (eds.), *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 138–147. Somerville, MA: Cascadia Proceedings Project. (<http://www.lingref.com/cpp/decemb/6/paper2241.pdf>)
- Stundžia, Bonifacas. 2016. Lithuanian. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, vol. 5, 3089–3106. Berlin, Boston: De Gruyter Mouton. (doi:10.1515/9783110424942-001)
- Ulvydas, Kazys (ed.). 1965. *Lietuvių kalbos gramatika*, vol. 1. Vilnius: Mintis.
- van Marle, Jaap. 1992. The relationship between morphological productivity and frequency: A comment on Baayen's performance-oriented conception of morphological productivity. In Booij, Geert & Van Marle, Jaap (eds.), *Yearbook of Morphology 1991*, 151–163. Dordrecht: Springer. (doi:10.1007/978-94-011-2516-1_9)
- Varvara, Rossella. 2019. Misurare la produttività morfologica: i nomi d'azione nell'italiano del ventesimo secolo. Bruno Moretti, Aline Kunz, Silvia Natale, Etna Krakenberger (eds.), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana (Berna, 6–8 settembre 2018)*, 187–201. Milano: Officinaventuno. (doi: 10.17469/O2102SLI000011)

Jurgis Pakerys, Agnė Navickaitė-Klišauskienė
Department of Baltic Studies
Vilnius University
Universiteto St 5
LT-01131 Vilnius
Lithuania
e-mail: jurgis.pakerys@flf.vu.lt, agne.navickaite@flf.vu.lt, flf.vu.lt

Virginijus Dadurkevičius
Institute of Digital Resources and Interdisciplinary Research
Vytautas Magnus University
V. Putvinskio St 23–216
LT-44243 Kaunas
Lithuania
e-mail: virginijus.dadurkevicius@vdu.lt