Jurgis PAKERYS, Agnė NAVICKAITĖ-KLIŠAUSKIENĖ
Vilnius University

Virginijus DADURKEVIČIUS
Vytautas Magnus University

PRODUCTIVITY OF DEVERBAL SUFFIXAL NOUNS IN THE JOINT CORPUS OF LITHUANIAN*

Abstract. We use the Joint Corpus of Lithuanian (1.3 billion tokens) to estimate the derivational productivity of the main suffixes of deverbal nouns in Lithuanian. We measure their realized, expanding, and potential productivity (B a a y e n 1992; 1993) and compare our rankings to the ones found in the main grammars and studies of neologisms. In general, the ordering of the suffixes according to their realized and expanding productivity is similar to the one found in earlier studies. The rankings according to potential productivity, however, in many cases proved to be less usable because suffixes with lower total frequencies of their formations appear to be overestimated when compared to suffixes having high total frequencies.

Keywords: Lithuanian; deverbal nouns; suffixal nouns; derivational productivity; realized productivity; expanding productivity; potential productivity; corpus linguistics.

1. Introduction
1.1. Previous research

The data on derivational productivity in Lithuanian are quite limited. Grammars and other works typically note that certain affixation patterns are more productive than others and list the affixes according to their productivity. The productivity itself is reflected in the number of synchronically transparent

formations, and as a result, some word-formation studies mention either exact or approximate numbers of formations, depending on the nature of their data sources. In addition, the use of a given affix for the formation of neologisms may be noted.

For example, in the description of action nominals, the main grammars list the most productive suffixes *-im-as* and *-ym-as* first, followed by *-t-is*, *-es-ys*, *-ul-ys*, etc.[1] Only for suffixes *-t-is* and *-es-ys* are approximate numbers of formations given (ca. 100) and it is noted that some of these derivations are neologisms. For the most productive suffixes, namely *-im-as* and *-ym-as*, however, no numbers are provided, which makes sense: action nominals in *-im-as* can be derived from any verb, with the exception of the class of verbs in inf. *-y-ti*, prs. 3 *-o*, and pst. 3 *-ė*, which derive their action nominals in *-ym-as*; therefore, counting all formations in *-im-as* and *-ym-as* is quite tedious. Finally, the derivatives with non-native suffixes of formations bearing neo-classical roots are listed separately and the most productive suffix, *-acij-a*, is found in ca. 200 derivatives (U r b u t i s 1965, 289–291, 302–303; 2006, 94–95, 99).

The issue of derivational productivity has been most recently addressed in the studies of derivatives found in the Database of Lithuanian Neologisms (M i l i ū n a i t ė, A l e k s a i t ė 2011), see M u r m u l a i t y t ė (2014; 2016; 2021), A l e k s a i t ė (2022), etc. New coinages found in poetry and fiction have also been shown to reflect some derivational trends, see V a s k e l i e n ė (2003; 2007; 2011; 2012). An overview of the main theoretical aspects of productivity before the advent of corpus-based research accompanied by some Lithuanian data can be found in U r b u t i s (1978, 262–274).

Corpora in general have rarely been used to assess word-formation processes in Lithuanian thus far. M i k e l i o n i e n ė (2000) discusses new lexical items in the media corpus of 1991–1996, and a significant part of the dataset is derivationally transparent words. As for the Corpus of Modern Lithuanian[2] and other sources, the formation of nouns and adjectives with neo-classical roots and affixes in Lithuanian is discussed in I n č i u r a i t ė - N o r e i k i e n ė (2015; 2017). K i e t y t ė (2023) discusses productivity of prefixes *nu-*, *pri-*, *į-*, and *iš-* in Lithuanian Web corpus (LithuanianWaC v2; SketchEngine).

---

[1] Suffixes are presented here and below followed by nom. sg. (or nom. pl. of pluralia tantum) endings that cumulatively mark case and number and also serve as indicators of declensions. Gen. sg. may be added when nom. sg. is not enough to identify the declension.

[2] 141M tokens, http://tekstynas.vdu.lt/tekstynas/index.jsp (last accessed on 12 May 2024).

1.2. Measuring derivational productivity in corpora

Quantitative measures of derivational productivity based on corpus data have gained popularity since their introduction in the early 1990s; see B a a y e n (1992; 1993) and general overviews in B a a y e n (2009), Z e l d e s (2012, 48–95), G a e t a, R i c c a (2015, 844–849), D a l, N a m e r (2016, 73–76). Three major types of productivity measures are distinguished: realized, expanding, and potential.

Realized productivity reflects the number of derivationally transparent words with a given affix (B a a y e n 2009, 901–902, 904–905). This estimate is the same as counting derivatives in the dictionary—one of the methods that was used to gauge productivity before the advent of corpora. The difference between the lemma list of a corpus and a typical dictionary is that the latter may exclude many derivational items for the sake of brevity (because of their regularity, rarity, register, etc.). The main drawback of the realized productivity measure is that it does not differentiate between well-established formations and recently derived ones (neologisms)—they are all counted together. To overcome this limitation, measures with hapax counts were suggested. Hapaxes occur only once in a certain corpus and, given the corpus is sufficiently large, hapax counts appear to be good indicators of the productivity of derivational processes. The motivation behind this is that the new formations initially have low frequency, and when one examines the list of hapaxes, a significant number of new formations can be found. All derivationally transparent hapaxes, of course, should not be expected to be neologisms and their counts are just indicators of the use of a given derivational process in the creation of new words (B a a y e n 2009, 902, 905–906).

Hapax counts are used in two measures of productivity: expanding and potential. In the case of expanding productivity, the hapaxes of derivationally-transparent items with a certain affix are divided by the total number of hapaxes of a given corpus, see B a a y e n (1993; 2009, 902, 905–906). This measure allows us to estimate the probability of encountering a new type (formation) in the set of hapaxes in the corpus. The absolute numbers of hapaxes with a given affix are also used as indicators of productivity in the same corpus, and we follow this practice in our study.

For the case of potential productivity, the number of hapaxes with a certain affix is divided by the total frequency of formations containing that affix (B a a y e n 1992; 2009, 902, 906). This estimates the probability of finding a new type (formation) with a given affix that has a frequency of

one as we sample the tokens of formations with that affix. This measure, however, sometimes produces counterintuitive rankings because the total frequencies of derived words produced by competing derivational processes may differ greatly, and comparatively low total frequencies may overestimate the productivity of the examined derivational processes (Van Marle 1992; Gaeta, Ricca 2006). This problem is solved when the potential productivity for competing derivational patterns is measured at the same total frequency values by applying a variable-corpus approach, see Gaeta, Ricca (2006) for the method and its application to Italian, or by using Large Number of Rare Events (LNRE) models (Evert, Baroni 2007), see application in, e.g., Štichauer (2009), Varvara (2019) on the data of Old Italian and Italian of the 21st century, respectively. In our study, we limit ourselves to the original estimation of potential productivity and leave the application of the abovementioned methods for future studies.

1.3. Our corpus and data extraction

For the purposes of our study, we needed an open access corpus of Lithuanian that would be sufficiently large and representative. The Joint Corpus of Lithuanian (JCL) was chosen as it contains 1.3 billion tokens and its lemma and token lists are openly available (Dadurkevičius 2020a; 2020b).[3] The representativeness and accessibility of this corpus, however, is not ideal. It comprises three different subcorpora: Lithuanian internet texts collected in 2014, legal texts of the Seimas (the Parliament) of the Republic of Lithuania (2011), and the Balanced Corpus of Modern Lithuanian (2008), see Table 1 and Dadurkevičius, Petrauskaitė (2020, 123–124). The latter two subcorpora are accessible online, while the first one has not been prepared for online access but was kindly provided to us by the compilers as a plain text file containing text lines (sentences) sorted alphabetically.

The lemmatizer used for the compilation of the lemma list of the JCL is based on the "Hunspell" platform and has a fixed dictionary and a set of inflectional rules (Dadurkevičius 2017). There is no morphological disambiguation of homographic forms and the lemmatizer returns all possible analyses of a given form. As a result, the total numbers of lemmas and their

---

[3] The token list is found in the file JCL_types_vs_DML6.txt from Dadurkevičius (2020b).

frequencies are sometimes inflated. We discuss this issue in Section 3.2 as it evidently affected the results of some agent noun suffixes where masculine and feminine nouns have a significant number of homographic forms; see also P a k e r y s et al. (2024).

T a b l e 1. **Subcorpora of the Joint Corpus of Lithuanian**

Subcorpus	Online access	Tokens
Lithuanian internet texts, Vil–nius 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

Although the dictionary of the lemmatizer is quite large, a significant number of derivatives were not recognized and remained unlemmatized. To compensate for this, we performed a semi–automatic lemmatization. First, we automatically filtered the corpus tokens according to the pattern SUFFIX + (all possible) ENDINGS and grouped the forms into potential lemmas; whenever possible, the base words were automatically added alongside the derivatives. We manually reviewed the resulting lemma lists and removed derivationally non-transparent formations and non-existent lemmas that were artificially created due to homographic forms or were just spelling errors. We also corrected or added the derivational bases when they were not indicated or were incorrectly added during the previous automatized step. Additional notes and examples regarding the problems of marking the formations as derivationally transparent are found in each section of the analyzed categories below.

We should also mention that although mostly native suffixes were investigated, a number of borrowed suffixes found in the nouns with neo-classical roots (also referred to as internationalisms) were included in our study. These words, be they nouns, adjectives, or verbs, are diachronically

independent borrowings, but there are many cases when they can be interpreted as bearing derivational relations synchronically. As borrowed elements, these lexemes were adapted in Lithuanian by the addition of affixes, such as the verbal suffix *–uo–* followed by inflectional affixes, e.g., *organiz-uo-ti* 'organize', or just the nominal endings, e.g., *organizacij-a* 'organization', *organizator-ius* (m.), *–ė* (f.) 'organizer'. When derivational relations in Lithuanian among such items are seen, they can be segmented and interpreted as being part of the derivational system, e.g., *organiz-acij-a*, *organiz-ator-ius*, *–ė* ← *organiz-uo-ti*.[4] The major grammars treat the suffixal internationalisms as derived when their potential bases co-exist synchronically (Urbutis 1965, 302, 336–337, 377, 393–394; 2006, 99, 112, 128, 132); see also (Murmulaitytė 2021, 50–52) for the derivational interpretation of new borrowings and formations in *–acij-a*.

In the following sections, we discuss our results of measuring the realized, expanding, and potential productivity of the following categories of Lithuanian suffixal deverbal nouns: action nominals (Section 2), agents (Section 3), instruments (Section 4), results and objects (Section 5), places (Section 6), and celebrations (Section 7).[5] We chose the most productive suffixes based on observations of the grammars, but in a few cases, the choice was limited by our ability to manually review large amounts of data (such exclusions will be noted in the corresponding sections below). Sometimes we also included additional suffixes when they were indicated as productive in the latest studies of neologisms.

2. Action nominals

2.1. We examined the following six suffixes of action nominals: five native (*–im-as*, *–ym-as*, *–yb-a*,[6] *–es-ys*, *–ul-ys*) and one borrowed (*–acij-a*), e.g., *aug-im-as* 'growth' ← *aug-ti* 'grow', *šild-ym-as* 'heating' ← *šild-y-ti* 'heat',[7] *gyn-yb-a* 'defense' ← *gin-ti* 'defend',[8] *skamb-es-ys* 'sound, ringing' ← *skamb-ė-*

---

[4]   The suffix *–uo–* of the base is omitted in the derivatives.

[5]   This is a cross-linguistically uncommon category; see motivation for distinguishing it in Lithuanian in Section 7.

[6]   This includes pluralia tantum in nom. pl. *–yb-os*.

[7]   The suffix *–y–* is only found in the infinitive stem and functions as an inflection class marker: inf. *šild-y-ti*, prs. 3 *šild-o*, pst. 3 *šild-ė*; stem *šild–* is taken as a base for *šild-ym-as*.

[8]   The past stem *gyn–* is taken as a base: inf. *gin-ti*, prs. 3 *gin-a*, pst. 3 *gyn-ė*.

ti 'sound, ring',[9] *svaig-ul-ys* 'dizziness' ← *svaig-ti* 'be dizzy', *registr-acij-a* 'registration' ← *registr-uo-ti* 'register'.[10] The grammars begin discussion of action nominals by first listing the most productive suffixes *–im-as* and *–ym-as*, which stand in complementary morphological distribution as mentioned in Section 1. They are not quantified, which is understandable given that both of them are of high productivity. Suffix *–t-is* follows next and, according to the grammars, has ca. 100 formations in standard Lithuanian, e.g., *tar-t-is* 'pronunciation' ← *tar-ti* 'pronounce'. We initially planned to measure the productivity of *–t-is*, but additional lemmatization produced too many potential lemmas that could not be manually reviewed given our time limits. After *–t-is*, suffix *–es-ys* is listed with ca. 100 formations, followed by *–ul-ys* (not quantified), and *–yb-a* (also not quantified, but it is mentioned that most of the formations used in standard Lithuanian are neologisms). Finally, suffix *–acij-a* is presented first among borrowed suffixes and is quantified as having ca. 200 formations; this means that it ranks third after *–im-as* and *–ym-as*. For further details on the suffixes mentioned above, see U r b u t i s (1965, 289–294, 302–303; 2006, 94–96, 99), S t u n d ž i a (2016, 3096).

Among current neologisms, suffix *–im-as* is the most productive, followed by much less numerous formations in *–acij-a* and *–t-is*; suffixes *–es-ys* and *–yb-a* are found only in occasional formations. It is notable that *–ym-as*, one of the most productive action nominal suffixes, is absent in this list (A l e k s a i t ė 2022, 45–47), see further notes in Section 2.2. Suffixes *–im-as*, *–ym-as*, *–t-is*, and *–es-ys* are also noted among formations coined by poetry and fiction authors (V a s k e l i e n ė 2017, 3–5).

Similar to other languages, derivatives with action nominal suffixes in Lithuanian not only are used in their basic function of denoting nominalized deverbal situations (actions/events), but also refer to results, objects, instruments, places, etc. (K o p t j e v s k a j a - T a m m 2015, 1196–1198; G a e t a 2015, 1213; M e l l o n i 2015, 1254; S z y m a n e k 2015, 1333). This complicates the annotation of the data because the same lemma may have several readings. For example, *įpjov-im-as* (← *įpjau-ti* 'incise'[11]) refers both

---

[9] The suffix *-ė(j)-* is found in the infinitive and the past stem and functions as an inflection class marker: inf. *skamb-ė-ti*, prs. 3 *skamb-a*, pst. 3 *skamb-ėj-o*; present stem *skamb-* is taken as a base for *skamb-es-ys*.

[10] The suffix *-uo-* of the base is not transferred to the derivative.

[11] This is based on the past stem *įpjov-*: inf. *įpjau-ti*, prs. 3 *įpjaun-a*, pst. 3 *įpjov-ė*.

to the action ('incising') and to the result ('cut, incision'), *gėr-im-as* (← *ger-ti* 'drink'[12]) refers to the action ('drinking') and to the object ('drink')[13], *vež-im-as* (← *vež-ti* 'transport') refers to the means ('carriage') and to the action ('transporting'), *įėj-im-as* (← *įei-ti* 'enter') refers to the place ('entrance') and to the action ('entering'), etc. We decided to count the noun as an action nominal if it could be used in reference to situations (actions) at least in some of its uses. This means that the total frequencies of some lemmas (as pure action nominals) are imprecise because they include all different uses of these lemmas. There is also an error margin of the annotators (us), who may have wrongly assumed or missed the possibility of a use of a given noun in reference to a situation. For example, *kalėj-im-as* (← *kal-ė-ti* 'be imprisoned') in the vast majority of its uses refers to the place ('prison'), but it also has a potential reading of a situation ('being imprisoned'); the latter use, however, could not be confirmed by corpus examples as the total number of tokens for the review was too high.[14] In the case of *organiz-acij-a* (← *organiz-uo-ti* 'organize'), which appears to be most frequently used to denote administrative structure, a reference to the action is also possible, which we were lucky to confirm during a partial review of the tokens, e.g., *renginys* […] *išsiskyrė savo tobula* **organizacija** 'the event […] was distinguished by its perfect organization' (subcorpus of Lithuanian internet texts). We also observed that stating the action/event function for formations based on prefixed (telic) verbs was difficult because such derivatives frequently denote results: consider *įpjov-im-as* above, which is based on the prefixed *į-pjau-ti*. In the case of nouns with the suffix *–yb-a*, we faced two difficulties: some formations could be derived both from nouns and from verbs (with a shortening of the base) and some apparently no longer refer to actions in modern Lithuanian and have lexicalized meanings. In the first case, we decided not to assume deverbal formation if the nominal base was available,

---

[12] This is based on the past stem *gėr-*: inf. *ger-ti*, prs. 3 *geri-a*, pst. 3 *gėr-ė*.

[13] This particular noun differentiates the action from the object by accentuation (*gḗr-im-as* 'drink' vs. *gėr-ìm-as* 'drinking'), but our corpus has no accentual information.

[14] The use of this noun as an action nominal is seen in the following example: *Mano* **kalėjimas** *tęsėsi šešius mėnesius* 'My imprisonment lasted six months' (https://prodeoet-patria.lt/files/html/Prunskis-prie-vilties-kryziaus.htm, text published in 1948, accessed on 12 April 2024). The notion of 'imprisonment' in Lithuanian is typically expressed by the action nominal derived from the transitive verb: *įkalin-ti* 'imprison' → *įkalin-im-as* 'imprisonment'.

e.g., *prek-yb-a* 'trade' ← *prek-ė* 'commodity' (alternative: ← *preki-au-ti* 'to trade'[15]). In the second case, we marked such formations as *tar-yb-a* 'council' (← *tar-ti(s)* 'consult, discuss'), *vald-yb-a* '(management) board' (← *vald-y-ti* 'manage, govern'), etc. as non-transparent (i.e., not having an action nominal reading) because such formations in modern Lithuanian are only used in specific lexicalized meanings.

Apart from the annotation problems just mentioned, we also carefully considered formations containing the negation prefix *ne-* so that we do not include derivatives where the action nominal suffix would be in the inner derivational cycle. As a result, we counted only the formations where the negation prefix belongs to the verbal base and is not added to the action nominal later and does not arise from spelling errors, e.g., *ne-sutar-im-as* 'disagreement' ← *ne-sutar-ti* 'disagree'. The cases where such an order of formation could not be confirmed were mostly spelling mistakes, i.e., the space after *ne* was erroneously omitted, e.g., *čia neįbrėžimas = čia ne įbrėžimas* 'it is not a scratch' (*įbrėž-im-as* 'scratch' ← *įbrėž-ti* 'to scratch').

2.2. The productivity measures of action nominal suffixes are presented in Table 2, where the suffixes are ranked according to the type counts.

Table 2. **Productivity measures of action nominal suffixes**

Suffix	Types	Hapaxes	Total frequency	Potential productivity ($*10^3$)
-im-as	15,274	2,855	41,351,162	0.0690
-ym-as	1,943	351	13,479,118	0.0260
-acij-a	966	69	3,731,629	0.0185
-es-ys	154	27	258,292	0.1045
-yb-a	116	29	2,618,160	0.0111
-ul-ys	89	23	43,058	0.5342

The ranking according to both realized and expanding productivity is virtually the same, except for the minor difference in the hapax counts of

---

[15] Suffix *-au-* of the base would be omitted in the derivative in this case (letter <i> in *preki-(au-ti)* marks palatalization before non-front vowels; palatalization before front vowels is automatic and not marked in writing: *prek-yb-a*, *prek-ė*).

–es-ys and *–yb-a* (27 vs. 29). The range of differences in total frequencies for the majority of suffixes is really significant and we leave the evaluation of the potential productivity at equal total frequencies for future studies.

The high productivity and regularity of the derivatives in *–im-as* and *–ym-as* is confirmed by the fact that their hapaxes and other low-frequency formations may include the aspectual prefixes *nebe–* 'not anymore' (discontinuative), rarely *tebe–* 'still' (continuative), and the restrictive prefix *te–* 'just (only)', e.g., *nebemylėj-im-as* 'not loving anymore' ← *nebe-mylė-ti* 'not love anymore', *nebelank-ym-as* 'not attending anymore' ← *nebelank-y-ti* 'not attend anymore' (75 formations), *tebegalioj-im-as* 'being still valid' ← *tebe-galio-ti* 'be still valid' (two formations[16]) *tesirūpin-im-as* 'caring only about smth.' ← *te-si-rūpin-ti* 'care only about smth.' (three formations[17]), etc.; see Arkadiev (2013) on action nominals with *nebe–*, Arkadiev (2011) on the (dis)continuative, and Arkadiev (2010) on the restrictive. Derivatives with these prefixes demonstrate a highly regular formation of action nominals in *–im-as* and *–ym-as* and reflect their verbal character, see Koptjevskaja-Tamm (2015, 1202–1204) on inheritance of verbal features in action nominals.

As mentioned earlier in Section 2.1, the Database of Lithuanian Neologisms surprisingly does not contain formations in *-ym-as*. A significant portion of hapaxes with *–ym-as* in our dataset are derived from verbs with the negative prefix *ne–* (ca. 95) and the majority of the other formations are based on prefixed verbs, which, in turn, may also have a reflexive (middle) affix (occurring also with *ne–*), e.g., *neišard-ym-as* ← 'not disassembling' ← *ne-iš-ard-y-ti* 'not to disassemble' (the base contains the prefixes *ne–* and *iš–*), *palank-ym-as* 'attending for some time' ← *pa-lank-y-ti* 'attend for some time' (the base contains the prefix *pa–*), *at-si-šald-ym-as* 'cooling down oneself' ← *at-si-šald-y-ti* 'cool down oneself' (the base contains the prefix *at–* and the reflexive affix *-si–*). It is true that they do not strike one's eye as intentionally coined and their regularity is one of the reasons why it is harder to spot such formations and include them in the database of neologisms. When a certain formation in *–ym-as* does look interesting, it is always its base that is newly

---

[16] Another formation is *tebesivadovav-im-as* 'being still guided by smth.' ← *tebe-si-vadovau-ti* 'be still guided by smth.' (based on the past stem: pst. 3 *tebesivadovav-o*).

[17] The other two formations are *tesurink-im-as* 'collection of just smth.', *tepripažin-im-as* 'recognition of just smth.' ← *te-surink-ti* 'collect just smth.', *te-pripažin-ti* 'recognize just smth.'.

coined or has a new sense, e.g., *atstabd-ym-as* 'cancelation of suspension' ← *atstabd-y-ti* 'cancel suspension' (the rare verb *at-stabdyti* is coined to mark the reverse action of the frequent *su-stabdyti* 'stop, suspend'), *įmąst-ym-as* 'transferring, putting (something somewhere) by reflecting, thinking' ← *įmąst-y-ti* 'transfer, put (something somewhere) by reflecting, thinking', etc.

It is interesting to note that the third suffix according to realized and expanding productivity is the borrowed *-acij-a*: this demonstrates that the share of derivationally interpretable borrowings is significant. Moreover, this suffix is also occasionally employed in true formations derived in Lithuanian (i.e., not historically borrowed), which proves that it functions as a live word-formation device, as already noted in Murmulaitytė (2021, 50–52), Aleksaitė (2022, 47), e.g., *persigrup-acij-a* 'regrouping' (hapax) ← *per-si-grup-uo-ti* 'regroup (itr.)' (the base has the Lithuanian prefix *per-*, the reflexive affix *-si-*, and the borrowed root *grup-*); the regular and commonly used formation from this base has the suffix *-im-as*: *persigrupav-im-as* 'regrouping'.

The observation made in the grammars that many formations in *-yb-a* are neologisms appears to no longer be valid. The formations that were considered neologisms in the 1960s, when the chapter on word formation (Urbutis 1965) was being prepared, are well-established lexemes today. It is true, however, that some new derivatives are still being coined; they typically refer to specialized fields and activities performed by professionals (as opposed to unmarked formations in *-im-as* and *-ym-as*), just as in the earlier period, see Urbutis (1965, 293; 2006, 95–96), e.g., *apskait-yb-a* 'accounting' ← *apskait-y-ti* 'account', *slaug-yb-a* 'nursing' ← *slaug-y-ti* 'nurse'. A few of them appear to be coined with a sense of irony, e.g., *(turto) plėš-yb-a* 'looting (of property) ← *plėš-ti* 'rob, loot'.

When evaluating the productivity of the abovementioned suffixes, we should bear in mind that the majority of them have limitations with regard to morphological structure and other features of the base (Urbutis 1965, 289–294, 302–303; 2006, 94–96, 99). As mentioned earlier, nominals in *-im-as* can be derived from any verb with the exception of the type with the inf. *-y-ti*, prs. 3 *-o*, pst. 3 *-ė*, which derives formations in *-ym-as*. The latter class appears to not be abundant if only non-prefixed and non-reflexive verbs are taken into account. However, as we noted above, the majority of the hapaxes in our dataset are derived from the prefixed, and also reflexive, verbs belonging to the said type, which expands the array of potential bases. So, the morphological class feeding the formations in *-ym-as* is not that small after

all, but it still cannot compete with a theoretically infinite set of the bases feeding the derivation of nominals in *–im-as*. The productivity of formations in *–acij-a* is constrained by their derivation from predominantly borrowed verbs with the suffix *–uo-ti*, but this limitation still allows the suffix to rank third. Of other less productive suffixes, formations in *–es-ys* are limited by their tendency to be derived from verbs with inf. *–ė-ti*, prs. 3 *–a*, pst. 3 *–ėj-o* that typically denote sound, visual states, etc. Formations in *–yb-a* and *–ul-ys* can be derived both from suffixal and non–suffixed verbs, but nominals in *–ul-ys* are limited by the tendency to derive them from bases denoting human psychological states and physiological processes. The formations in *–yb-a*, as noted earlier, are also limited by their tendency to denote specialized activities and fields and not just pure actions.

3. Agent nouns

3.1. Our survey includes six suffixes of deverbal agent nouns: five native (*–toj–*, *–ėj–*, *–ik–*, *–ėl–*, *–ūn–*) and one borrowed (*–ator–*). Agent formations in Lithuanian are either masculine or feminine, and their gender relates to specific declensions: masculine agents are in the nom. sg. *–as*, *–is* (gen. sg. *–io*), *–ius* and feminine ones are in the nom. sg. *–a*, *–ė*, e.g., *vairuo-toj-as* m., *vairuo-toj-a* f. 'driver' ← *vairuo-ti* 'drive', *pirk-ėj-as* m., *pirk-ėj-a* f. 'buyer' ← *pirk-ti* 'buy', *plauk-ik-as* m., *plauk-ik-ė* f. 'swimmer' ← *plauk-ti* 'swim', *nusikalt-ėl-is* m., *nusikalt-ėl-ė* f. 'criminal' ← *nusikals-ti* 'commit a crime',[18] *klaj-ūn-as* m., *klaj-ūn-ė* f. 'wanderer' ← *klaj-o-ti* 'wander',[19] *organiz-ator-ius* m.,[20] *organiz-ator-ė* f. 'organizer' ← *organiz-uo-ti* 'organize'. The traditional approach of Lithuanian grammars and other studies is to discuss masculine and feminine agent nouns together (and sometimes explicitly qualify them as substantiva mobilia), but we treated such formations as distinct uses of the corresponding suffixes, which allowed us to reveal important productivity differences between the genders, see the notes in Section 3.2; that masculine

---

[18] The past stem *nusikalt–* serves as a base: inf. *nusikals-ti*, prs. 3 *nusikalst-a*, pst. 3 *nusikalt-o*.

[19] The suffix *–o–* of the base is omitted in the derivative.

[20] The letter <i> in the nom. sg. *–ius* indicates the palatalization of the preceding consonant. A phonologically and morphologically preferable segmentation would be *–atori-us*, but for the sake of clarity (graphic uniformity of the suffix), we use the segmentation *–ator-ius*. The same applies to the segmentation of the gen. sg. *–io* mentioned earlier.

and feminine agents are independently derived from the base verbs is noted in, e.g., Urbutis (1965, 417).

The suffix *-toj-* is characterized as the most productive of all the agents proper in the grammars, while *-ėl-* is seen as the most productive suffix of the bearers of verbal features and is listed second (see more on this semantic subcategory of agents below), but no quantitative data are provided. The suffix *-ėj-* comes third, followed by a less productive *-ik-*; the grammars estimate that formations with *-ik-* are less numerous by approximately one-third compared to the ones with *-ėj-*. The suffix *-l-* (*-l-ys, -ė*) comes next (e.g., *ved-l-ys, -ė* 'guide' ← *ves-ti* 'lead'[21]) without any quantification; we initially intended to evaluate its productivity, but our method of additional lemmatization produced too many potential lemmas that we were unable to review manually given our time limits. Then, the suffix *-ūn-* follows and is specified as not rare, but produces a few neologisms (the approximate number of derivatives is not given). Finally, the borrowed suffix *-ator-* is characterized as found in ca. 50 formations and comes first in the list of suffixes found in internationalisms. See more details on the agent formation with these suffixes in Urbutis (1965, 317–323, 336–337; 2006, 104–107, 112), Stundžia (2016, 3097).

Studies of recent neologisms note the outstanding productivity of *-toj-* formations and *-ėl-* is also seen as productive to some extent, but the rest of the suffixes of our survey are found in only a few derivatives or are not discussed (*-ator-*), see Murmulaitytė (2016; 2021, 149–159), Aleksaitė (2022, 57–63). Some formations in *-toj-*, *-ėl-*, and *-ėj-* are also found among new coinages by Lithuanian poetry and fiction authors (Vaskelienė 2017, 5).

As for semantics, it is well known that the term "agent noun" covers a broad range of formations where some of the derivatives are not agents at all, i.e., not the initiators or the causers of events (Rainer 2015, 1305–1307), e.g., *gyven-toj-as, -a* 'inhabitant' ← *gyven-ti* 'live', *paveldė-toj-as, -a* 'inheritor' ← *paveldė-ti* 'inherit', etc. This is especially true of formations in *-ėl-*, which can be paraphrased by resultative constructions with participles, i.e., 'the one who is V-ed' and not 'the one who V-s', e.g., *numir-ėl-is, -ė* 'the one who is dead' ← *numir-ti* 'die', *pamiš-ėl-is, -ė* 'the one who is gone mad' ← *pamiš-ti* 'go mad'. Lithuanian word-formation studies usually differentiate between the agents proper, or the conscious performers of the

---

[21] The stem *ved-* serves as a base: inf. *ves-ti*, prs. 3 *ved-a*, pst. 3 *ved-ė*.

actions, professionals, etc. (typically in *-toj-*, *-ėj-*, *-ik-*, *-ator-* in our set of suffixes) and the animate and inanimate entities characterized by their salient properties denoted by the base verbs; the latter formations are referred to as "the bearers of verbal features" (formed with the suffixes *-ėl-* and *-ūn-* in our sample). The distinction between the agents proper and the bearers of verbal features is not always clear-cut, but it is notable that only the latter show a tendency to have a pejorative character and/or indicate a poor state of the person; consider the two examples with *-ėl-* above (U r b u t i s 1965, 318, 322; 2006, 105, 107). These semantic subcategories in Lithuanian word-formation studies are always treated together, see U r b u t i s (1965, 317; 2006, 104), S t u n d ž i a (2016, 3097). Finally, we should note that derivatives in *-toj-*, *-ėj-*, *-ator-* (very rarely in *-ik-* and *-ūn-*) may also denote instruments and means, reflecting a well-known multifunctionality of such affixes (R a i n e r 2015, 1308). The data presented in Section 3.2 below reflect the use of agent formations, while instrument nouns formed with prototypical (or frequent[22]) agent suffixes are discussed later in Section 3.3.

The relations between the derivatives and the bases in our data were in most cases transparent, except for some formations where the semantic motivation was less evident, e.g., *(teismo) tar-ėj-as*, *-a* '(court) counsellor' ← *tar-ti* 'say' (cf. also prefixed *pa-tar-ti* 'advice'), or the bases are very rare in current use, e.g., *pribuv-ėj-a* f. (rarely *-as* m.) '(traditional) midwife' ← *pribū-ti* 'be present (for longer time)'.[23] We should also note that we limited the analysis of formations in *-ėl-is*, *-ė* to only the cases where potential bases were indicated during our semi-automatic lemmatization process. A full review of all potential lemmas in this case was impossible due to time restrictions, and we hope to get updated results in the future after we have analyzed and separated the denominal diminutives in *-ėl-is*, *-ė* from the deverbal formations with the same suffix.

3.2. Productivity measures of the agent noun suffixes are presented in Table 3. The suffixes are listed according to the sum of the type counts of the masculine and feminine formations with the same suffix. Feminine formations are listed immediately after the corresponding masculine formations for convenience of comparison.

---

[22] The case of *-ator-*.

[23] The past stem *pribuv-* is taken as a base: inf. *pribū-ti*, prs. 3 *pribūn-a*, pst. 3 *pribuv-o*.

Table 3. **Productivity measures of agent noun suffixes**

Suffix	Types	Hapaxes	Total frequency	Potential productivity (*10³)
–toj–as m.	2,402	447	6,520,043	0.0686
–toj–a f.	2,279	69	4,710,297	0.0146
–ėj–as m.	618	100	3,899,405	0.0256
–ėj–a f.	620	25	2,973,383	0.0084
–ėl–is m.	657	186	254,357	0.7313
–ėl–ė f.	316	52	95,173	0.5464
–ik–as m.	253	62	205,327	0.3020
–ik–ė f.	89	23	21,306	1.0795
–ator–ius m.	164	26	513,296	0.0507
–ator–ė f.	138	6	157,079	0.0382
–ūn–as m.	75	22	16,512	1,3324
–ūn–ė f.	34	9	1,005	8.9552

The ranking according to realized productivity, if masculine and feminine formations are summed together, generally confirms the observations found in earlier studies. A notable difference is that *–ėl–* comes before *–ėj–* in the grammars, while our data differ: the total number of *–ėj–* formations supersedes that of *–ėl–*. We should admit that we are unsure about this finding because the number of feminine formations in *–ėj–* in our counts seems to be inflated due to abundant homographic forms (see more details below, especially of note are manually reviewed hapax counts that result in *–ėl–* formations ranking second).

An important and new aspect of our study is that the feminine formations are shown to be clearly less numerous than the masculine ones. This difference is likely due to a wider use of masculine formations as generic terms (consider also the hapax counts and their discussion below), but further research is needed to establish the extent of the use of generic masculine agent formations. We are also looking forward to having more precise type counts when the problem of morphological ambiguity during the lemmatization

stage is solved to an acceptable degree (see also the notes and results of the manual disambiguation of hapaxes below).

As mentioned earlier, the grammars cautiously observe that agent nouns in *–ik–* are less numerous by around one-third compared to those in *–ėj–* (U r b u t i s 1965, 320; 2006, 106). In order to evaluate our data in this respect, we disregarded the distinction between masculine and feminine nouns (as the grammars do) by removing the endings and deleting duplicate stems. This resulted in 687 unique stems in *–ėj–* and 263 stems in *–ik–*. The difference between *–ėj–* and *–ik–* types in our corpus is much more significant than suggested in the grammars: there are 2.61 times more formations in *–ėj–*. It is likely that this difference results from examining corpus *versus* dictionary data, and we believe that the evaluation presented in the grammars is based on certain dictionary counts done in the 1960s.[24] Finally, our data allowed formations in the borrowed *–ator–* to be ranked (as fifth), while the grammars lacked quantitative data for many agent noun suffixes and, as a result, *–ator–* could not be ranked.

The ranking according to expanding productivity is similar. We noted earlier that the type counts of *–ėj–* might be incorrect due to lemmatization errors. Hapax counts, however, were manually reviewed (see below) and they demonstrate that *–ėl–* ranks second, followed by *–ėj–* (recall that *–ėl–* is semantically different from the majority of other suffixes as it is used to denote bearers of the verbal features rather than true agents). Studies of neologisms note the scarcity of formations in *–ėj–* (M u r m u l a i t y t ė 2016, 15–16; 2021, 157; A l e k s a i t ė 2022, 62), but we guess that as the Database of Lithuanian Neologisms grows, *–ėj–* (and maybe also *–ik–*) formations will be more noticeable; consider some hapaxes from our corpus: *atsivež-ėj-as* m. 'the one who brings smth. for oneself' ← *atsivež-ti* 'bring for oneself (by transportation)', *praloš-ėj-as* m. 'loser (of the game) ← *praloš-ti* 'lose (the game)', *atsiėm-ik-as* m. 'the one who takes back smth.' ← *atsiim-ti* 'take back',[25] etc. Of note is also the difference between the hapax counts of the native suffix *–ik–* (85 in total) and its lower neighbor in the ranking—suffix *–ator–* of the borrowings (32 in total). The native suffix clearly prevails, but it is still

---

[24] The data of the following dictionary could have been used: Jonas K r u o p a s et al. (eds.), *Dabartinės lietuvių kalbos žodynas*, Vilnius: Valstybinė politinės ir mokslinės literatūros leidykla, 1954.

[25] The past stem *atsiėm-* is used as a base: inf. *atsiim-ti*, prs. 3 *atsiim-a*, pst. 3 *atsiėm-ė*.

worth noting that some formations in *-ator-* are actually not borrowed but are formed independently in Lithuanian, which shows that *-ator-* functions as a true word-formation device giving rise to such formations as, e.g., *film-ator-ius* m. 'the one who films' ← *film-uo-ti* 'film', *kopij-ator-ė* f. 'imitator (i.e., the one who copies smth.)' ← *kopij-uo-ti* 'copy'[26] found alongside regular formations in *-toj-*: *filmuo-toj-as* m., *kopijuo-toj-a* f.

Now let us come back to the type counts. We noticed that our lemmatizer, which has no disambiguation module, produced a significant number of potential lemmas that were based on homographic forms of masculine and feminine nouns, especially for the suffixes *-toj-* and *-ėj-*, which have a large number of homographic cells in their case paradigms. For example, the corpus contains the hapax form *apčiupinėtojas*, which can be either the nom. sg. of the masculine agent noun or the acc. pl. of the feminine agent noun. The lemmatizer could not disambiguate and thus returned two potential lemmas: *apčiupinė-toj-as* m. and *apčiupinė-toj-a* f. 'the one who feels/checks by touching' ← *apčiupinė-ti* 'feel/check by touching'. We were unable to resolve the issue of morphological ambiguity for all types as this would require a very time-consuming manual review of all the homographic forms. Instead, we decided to review and disambiguate all the hapaxes. The results presented in Table 3 above already reflect the counts of manually disambiguated hapaxes, i.e., when the context showed that the masculine noun was used, the potential feminine lemma was deleted from that list, or vice versa. There were cases when we could not make strict judgements and decided to count them separately, i.e., as potentially feminine or masculine forms; many such contexts appear to reflect the generic use of masculine nouns, but we used a conservative approach and marked them as ambiguous. The numbers of hapaxes that were counted as ambiguous, i.e., masculine or feminine (and not included in the hapax counts in Table 3 above), are as follows: *-toj-* 129, *-ėj-* 30, *-ėl-* 25, *-ator-* 9, *-ik-* 4, *-ūn-* 0. The corrected hapax counts show that the formation of masculine agent nouns significantly surpasses that of feminine nouns. Thus, the rather small difference between the type counts of the masculine and feminine formations in *-toj-* and *-ėj-* noted earlier is most likely due to inflated numbers of feminine formations arising from homographic forms. Future studies employing lemmatizers with good disambiguation capabilities should shed more light on this issue. A

---

[26] The suffix *-uo-* of the base is omitted in the derivatives.

more detailed discussion of the lemmatization problems of Lithuanian agent nouns can be found in P a k e r y s et al. (2024).

Measures of potential productivity are not easily comparable in many cases due to differences in total frequencies. In addition, we should bear in mind that because of unresolved morphologically ambiguous forms, total frequencies are also imprecise, especially in the case of the suffixes *-toj-* and *-ėj-*. For example, the ranking of *-ėj-as* m. after *-toj-as* m. and *-ėj-a* f. after *-toj-a* f. appears to be credible, but still needs to be reevaluated after lemmatization with a good level of morphological disambiguation and when potential productivity is measured at the same total frequency values. Another interesting fact is that the suffix *-ėl-is* m. clearly supersedes *-ik-as* m.: this also appears to be a credible finding given that their total frequencies are quite similar (254,357 and 205,327).

Finally, we should note that many suffixes under consideration have a morphological distribution with respect to their bases, which limits their productivity: *-toj-* is added to suffixal bases, while *-ėj-* and *-ik-* are added to non-suffixal bases; very few exceptions exist for all three suffixes (U r b u t i s 1965, 318–321; 2006, 105–106). Therefore, a competition under equal conditions, i.e., with the same arrays of potential bases, is seen only for *-ėj-* and *-ik-*, where *-ėj-* wins. Furthermore, the outstanding productivity of *-toj-* can be explained by the fact that the suffixed verbs themselves are a productive type and outnumber the non-suffixed ones (J a š i n s k a i t ė 1971, 247), which feed the derivation in *-ėj-* and *-ik-*. The array of bases for *-ator-* formations is also limited: this suffix derives agents only from verbs that end in the suffix *-uo-*, which mostly have neo-classical roots. The suffix *-ėl-* is usually added to non-suffixed verbs and is directly comparable to *-ėj-* and *-ik-*; one should remember, however, that the semantics of the derivatives differs: *-ėl-* produces the bearers of the verbal features, while *-ėj-* and *-ik-* derive agents (U r b u t i s 1965, 319; 2006, 105). Finally, the suffix *-ūn-* (which ranks last in our set) is interesting in that it is not actually limited by the morphological types of the bases since it is added to both suffixed and non-suffixed stems (U r b u t i s 1965, 322; 2006, 107).

3.3. As mentioned in Section 3.1, formations in *-toj-*, *-ėj-*, *-ator-*, and very rarely also in *-ik-* and *-ūn-* may denote instruments and means. The use of *-toj-* and *-ėj-* in this function is proscribed in standard Lithuanian and not mentioned in the grammars (see some references at the end of this section). Sometimes, the same noun can be used in reference to agents and instruments

(means), e.g., *nuėm-ėj-as* m. 'remover', as in the phrases *derliaus nuėmėjas* '(lit.) remover of the harvest = the worker who harvests' and *lako nuėmėjas* 'remover of the polish = substance that removes polish' ← *nuim-ti* 'remove'.[27] Noticing such cases is not always easy, especially when a typical instrument noun is only occasionally used in reference to an agent. We checked all the contexts of the hapaxes, but for more frequent nouns, we relied on our expertise as native speakers; we also marked the nouns as instruments only in the case of more or less lexicalized instrument designations. This means that there could be more instrument nouns that are occasionally used in reference to agents and vice versa (some agent nouns that occasionally refer to instruments). Moreover, it is not easy to draw a clear line between an instrument and a characterization of a certain object, substance, etc. performing an action (= non-animate agent). We tried to qualify the nouns as instruments only in the cases in which they were more or less conventionalized. Prototypical instruments derived with our set of the suffixes are of masculine gender, but occasional uses of feminine nouns with the corresponding function are also found when the object characterized by the derivative is of feminine gender, e.g., *stimuli-ator-ė* f. 'stimulant' (← *stimuli-uo-ti* 'stimulate') characterizing *imbiero šaknis* f. 'ginger root' as a 'stimulant': *Imbiero **šaknis** pasaulyje žinoma kaip jausmų **stimuliatorė*** 'Ginger root is known worldwide as a stimulant of the feelings' (subcorpus of Lithuanian internet texts).

Table 3 in Section 3.2 above presents the counts of all nouns that are used as agents only. Some of them are also used to denote instruments (means), an aspect seen in Table 4 below; consider the second column: types (agent and instrument). The nouns that are used only as instruments in our corpus were excluded from Table 3 in Section 3.2 but are presented in Table 4 below (types and hapaxes are in the third and fourth columns).

The suffix *-ator-ius* (found in the internationalisms) stands out here as frequently found among both agent and instrument formations. The instrument function of this suffix (alongside its agentive function) is acknowledged in the grammars (U r b u t i s 1965, 393–394; 2006, 132). Our data also confirm the fact that instruments in *-ik-as* and *-ūn-as* are very rare. The suffix *-ik-as* is listed in the grammars among the instrument affixes of low productivity, while *-ūn-as* is characterized as found only in occasional instrument formations (U r b u t i s 1965, 389–391; 2006, 131–132). The

---

[27] The past stem *nuėm-* is used as a base: inf. *nuim-ti*, prs. 3 *nuim-a*, pst. 3 *nuėm-ė*.

Table 4. **Types and hapaxes of instrument nouns with suffixes that also derive agent nouns (ranked according to the types referring to instruments only)**

Suffix	Types (agent and instrument)	Types (instrument)	Hapaxes (instrument)
–ator–ius	31	177	16
–ėj–as	30	39	13
–toj–as	21	34	16
–ik–as	0	3	1
–ūn–as	0	1	0

suffixes *–toj-as* and *–ėj-as*, however, are not discussed in the grammars due to the prescriptive tradition that does not acknowledge the formation of such instrument nouns in standard Lithuanian, see, e.g., Paulauskienė et al. (1976, 70), Pupkis (1980, 118).[28] The need to proscribe them, of course, shows a certain degree of productivity in everyday use, which is reflected in our corpus data. These suffixes, however, are not as productive as the prototypical instrument suffixes *–tuv-as* and *–ik-as* discussed in Section 4. Further study is needed to determine to what extent the types of the texts influence the productivity of instruments in *–toj-as* and *–ėj-as* because they appear to also be dependent on editorial practices.

4. Instrument nouns

4.1. Six suffixes of instrument nouns are included in our survey. The majority of them are native (*–tuv-as*, *–ikl-is*, *–ykl-ė*, *–tuk-as*, *–tuv-ė*) and one (*–ator-ius*) is found in internationalisms, e.g., *šaldy-tuv-as* 'refrigerator' ← *šaldy-ti* 'freeze', *jung-ikl-is* 'switch' ← *jung-ti* 'switch', *vir-ykl-ė* 'stove' ← *vir-ti* 'cook, boil', *pieš-tuk-as* 'pencil' ← *pieš-ti* 'draw', *kep-tuv-ė* 'pan' ← *kep-ti* 'bake', *akumuli-ator-ius* 'accumulator' ← *akumuli-uo-ti* 'accumulate'.[29] The

---

[28] See also "The list of language errors" (The State Commission of the Lithuanian Language, 1997): https://www.vlkk.lt/aktualiausios-temos/didziosios-klaidos/zodziu-sandaros (Section 2.2.2); last accessed on 12 May 2024.

[29] The suffix *–uo-* of the base is omitted in the derivative.

inflection class selected by the suffix sometimes is a non-trivial indicator of the derivational category, consider *–ykl-ė* 'instrument' (declension *-ė*) vs. *–ykl-a* 'place' (declension *-a*); a less strict differentiation is seen in *–tuv-as* 'instrument' (declension *-as*) vs. *–tuv-ė* 'instrument' or 'place' (declension *-ė*).

Grammars characterize the instrument formations as generally not very productive. The most productive suffixes are *–tuv-as* (ca. 200 formations) and *–ikl-is* (ca. 100 formations), followed by *–t-as* (ca. 50 formations), *–tuk-as* (ca. 50 formations), *–al-as* (ca. 50 formations), and *–ykl-ė*, and *–tuv-ė* (the number of formations is not indicated for the latter two suffixes); *–ator-ius* is not quantified but is listed first among the suffixes found in the internationalisms (U r b u t i s 1965, 381–386, 393–394; 2006, 129–130, 132; S t u n d ž i a 2016, 3097). Our survey includes the most productive suffixes (*–tuv-as*, *–ikl-is*), one borrowed suffix (*–ator-ius*), and two additional suffixes that show some productivity in the formation of neologisms (*–tuk-as*, *–ykl-ė*, see M u r m u l a i t y t ė 2021, 135) in contrast to other suffixes mentioned above, i.e., *–t-as*, *–al-as*, which were not included in our study. The suffix *–tuv-ė* was also initially not included, but the examination of place nouns with this suffix revealed that its use in the formation of instrument nouns is quite noticeable.

As for the semantic subtypes of instruments, the borrowed *–ator-ius* is predominant in designations of modern instruments because the majority of internationalisms refer to relatively new inventions. The native suffixes, however, are also found in designations of modern instruments, consider the examples above; for the distinction of 'traditional' vs. 'modern' instrument nouns, see R a i n e r (2015, 1311). The suffix *–tuk-as* is special in that it has a diminutive connotation and is used to refer to small instruments, consider 'pencil' above (U r b u t i s 1965, 384; 2006, 129). Our data show that the instrument noun suffixes, especially *–tuv-as*, can occasionally be used to derive agent nouns with an ironic or pejorative sense, e.g., *dėsty-tuv-as* m. 'university teacher, professor' (jokingly) alongside regular and semantically neutral *dėsty-toj-as* m. ← *dėst-y-ti* 'teach (at the university)'. The suffix *–ator-ius* is also used to derive agent nouns (see Section 3.3), while the suffix *–tuv-ė* is found in deverbal place nouns (see Section 6); for these particular types of suffix multifunctionality, see R a i n e r (2015, 1308), S z y m a n e k (2015, 1332–1334).

The relations between derivatives and their bases in most of the cases were transparent, except for some formations when the base was rare or a specific sense of the base was rather marginal. We followed a liberal approach and marked them as transparent, e.g., *myg-tuk-as* 'button' ← *myg-ti* 'press'

(the base verb is rather rare) and *svarst–ykl–ės* 'scales' ← *svarst–y–ti* 'weigh (repeatedly)' (the specific sense of the base is very rare).

4.2. Productivity measures of the instrument suffixes are presented in Table 5 where the suffixes are arranged by the type counts.

Table 5. **Productivity measures of instrument noun suffixes**

Suffix	Types	Hapaxes	Total frequency	Potential productivity (*10^3)
–tuv–as	488	87	414,902	0.2097
–ikl–is	293	54	232,475	0.2323
–ator–ius	177	16	122,040	0.1311
–tuk–as	116	23	67,972	0.3384
–tuv–ė	123	20	25,042	0.7987
–ykl–ė	102	21	57,039	0.3682

The ranking of the suffixes according to both realized and expanding productivity is similar and generally reflects the order found in the grammars. A new finding is that, according to the type counts, the instruments in *–ator–ius* rank third after two of the most productive native suffixes, revealing that such formations (diachronically mostly borrowings) play an important role in the derivationally transparent part of the lexicon. The number of hapaxes in *–ator–ius*, however, is rather low, ranking this suffix as last in our set. This demonstrates that new formations (or, in many cases, derivationally segmentable new borrowings) are comparatively rare. For an example of a true formation, i.e., most likely derived in Lithuanian and not borrowed, consider *lituaniz–ator–ius* 'Lithuanizer (an application that adds Lithuanian fonts and keyboard)' ← *lituaniz–uo–ti* 'Lithuanize, make Lithuanian'.[30] The higher potential productivity of *–tuk–as*, *–tuv–ė*, and *–ykl–ė* can be disregarded for the time being due to their rather low total frequencies compared to the first three suffixes; *–tuv–ė* also stands out in its significantly lower total frequency compared to *–tuk–as* and *–ykl–ė*.

Studies of recent neologisms note that formations in *–ikl–is* rank first (Murmulaitytė 2021, 135–143; Aleksaitė 2022, 76–78). Our data

---

[30] The suffix *–uo–* of the base is omitted in the derivative.

differ, however, in that the suffix *-tuv-as* has a high number of hapaxes, whereas the Database of Lithuanian Neologisms has registered rather low numbers of new formations and, as a result, *-tuv-as* ranks only third or fourth among other suffixes; the suffix *-tuv-ė* is found only in occasional formations (M u r m u l a i t y t ė 2021, 135–137; A l e k s a i t ė 2022, 77–79). We believe that more formations in *-tuv-as* might be included in the database in the future; for some examples of hapaxes from our dataset, consider *atsaky-tuv-as* '(automatic e-mail) responder' ← *atsak-y-ti* 'respond', *garbano-tuv-as* '(hair) curler' ← *garbano-ti* 'curl (hair)', *mulčiuo-tuv-as* 'mulcher' ← *mulči-uo-ti* 'make into mulch'.

One should also bear in mind that some instrument suffixes have restrictions with regard to their bases: instruments in *-ator-ius* are derived only from verbs that end in the suffix *-uo-*, which mostly have neo-classical roots; formations in *-tuk-as* are derived only from non-suffixed verbs. The rest of the suffixes do not have limitations with regard to their bases: this evidently helps the suffixes *-tuv-as* and *-ikl-is*, but not, e.g., the suffix *-ykl-ė*, which has low productivity.

5. Result and object nouns

5.1. Three suffixes of result and object nouns were selected for our study: two native ones, which are listed in the grammars as the most productive (*-in-ys*, ca. 200 formations; *-al-as*, ca. 100 formations) and one borrowed suffix (*-at-as*, quantitively not characterized in the grammars), e.g., *rink-in-ys* 'set' (result) ← *rink-ti* 'collect', *siunt-in-ys* 'parcel' (object) ← *siųs-ti* 'send',[31] *tirp-al-as* 'solution' (result) ← *tirp-in-ti* 'dissolve', *les-al-as* 'bird feed' (object) ← *les-ti* 'feed (itr.; about birds)', *falsifik-at-as* 'forgery' (result) ← *falsifik-uo-ti* 'falsify', *ekspon-at-as* '(an) exhibit' (object) ← *ekspon-uo-ti* '(to) exhibit, display'[32] (U r b u t i s 1965, 368–370, 377; 2006, 125, 128; S t u n d ž i a 2016, 3096).

With respect to function, our suffixes are used to derive result and object nouns and their interpretation depends on the semantics of the base verb, for example, creation and modification verbs are likely to produce result nouns, consider examples above; see also M e l l o n i (2015, 1264–1265).

---

[31] The past stem *siunt-* is used as a base: inf. *siųs-ti*, prs. 3 *siunči-a*, pst. 3 *siunt-ė*.

[32] The suffixes *-in-* and *-uo-* of the bases are omitted in *tirp-al-as*, *ekspon-at-as*, and *falsifik-at-as*.

Consequently, result and object nouns are discussed as a single category (U r b u t i s 1965, 367–381; 2006, 125–128). As mentioned in Section 2, some action nominals also bear the semantics of result nouns, but here we focus only on the specialized suffixes that prototypically derive result or object nouns, similarly to, e.g., Slavic languages (M e l l o n i 2015, 1254). We also do not separate human affected participants into a special category of patient nouns (M ü h l e i s e n 2015), consider some formations found with *-in-ys* in our dataset, e.g., *kank-in-ys* 'martyr' ← *kank-in-ti* 'torture', *mok-in-ys* 'pupil' ← *mok-y-ti* 'teach'.[33] Our interpretation here diverges from the grammars where such formations (i.e., referring to humans) are listed as 'bearers of verbal features', see U r b u t i s (1965, 328; 2006, 109). Finally, it should be mentioned that some derivatives in *-al-as* are noted for their pejorative semantics (U r b u t i s 1965, 370; 2006, 125), see examples in the following section.

5.2. Productivity data of the result and object noun suffixes are presented in Table 6; the ranking of the suffixes is according to the type counts.

T a b l e 6. **Productivity measures of result and object nouns**

Suffix	Types	Hapaxes	Total frequency	Potential productivity (*10³)
-in-ys	230	16	3,617,231	0.0044
-al-as	124	31	324,873	0.0954
-at-as	72	5	181,609	0.0275

The ranking of the first two suffixes according to realized productivity corresponds to that suggested by the grammars; our input here is the added quantification of the borrowed *-at-as*. Expanding productivity, however, shows an interesting result by ranking the formations in *-al-as* first. Perhaps this can be explained by the need of the speakers to express the pejorative semantics when coining ad hoc formations, as seen in the following examples: *Gal tas **mezgalas** kada nors ir buvo šalikas?* 'Maybe that **miserable knit** was once a scarf?' (Balanced Corpus of Modern Lithuanian; *mezg-al-as* 'miserable

---

[33] The suffix *-in-* is omitted in the derivative and the suffix *-y-* is found only in the infinitive stem (functioning as an inflection class marker).

knit' ← *megz-ti* 'to knit'[34]); *Daro vien tam*, *kad daryti – tai* **daralas** *ir gaunasi* 'They do it just for the sake of doing, so they [expectedly] get a bad result' (Lithuanian internet text subcorpus; *dar-al-as* 'poor result of doing smth.' ← *dar-y-ti* 'do'[35]); cf. non-pejorative result nouns with the suffix *-in-ys*: *mezg-in-ys* '(a) knit', *dar-in-ys* 'formation'. The Database of Lithuanian Neologisms contains only occasional formations in *-in-ys* (2 nouns) and *-al-as* (1 noun) (Aleksaitė 2022, 74–75), but as our data show, some formations with these suffixes are likely to be registered in the future, consider the following hapaxes from the corpus: *pratęs-in-ys* 'continuation' ← *pratęs-ti* 'continue (tr.)', *sutelk-in-ys* 'concentration' ← *sutelk-ti* 'concentrate (tr.)', *žaid-al-as* 'game' ← *žais-ti* (past stem *žaid-ė*) 'play', etc.

6. Place nouns

6.1. We surveyed the two deverbal place noun suffixes listed as the most productive ones in the grammars: *-ykl-a* (ca. 100 formations) and *-tuv-ė* (more than a dozen formations), e.g., *kep-ykl-a* 'bakery' ← *kep-ti* 'bake', *dirb-tuv-ė* 'workshop' ← *dirb-ti* 'work' (Urbutis 1965, 399, 401; 2006, 134–135; Stundžia 2016, 3097). As mentioned earlier in Section 4.1, the declensions selected by the suffixes sometimes help differentiate categories of place vs. instrument, which is especially evident in the case of *-ykl-*: *-ykl-ė* 'instrument' vs. *-ykl-a* 'place' (a few ambiguous examples can be found); not so, however, in the case of *-tuv-ė*: mostly 'instrument', but also 'place', see Section 6.2.

The formations in our dataset were mostly transparent, except for some cases of an ambiguous interpretation between 'instrument' and 'place', e.g., *kar-tuv-ės* 'gallows' (plurale tantum in the nom. pl. *-ės*) ← *kar-ti* 'hang'; we chose 'instrument', as in (Urbutis 1965, 386). The multifunctionality of suffixes denoting 'instrument' and 'place (and other categories) is well known, see Szymanek (2015, 1332–1334). As for some other categories already discussed above, we noted that the semantic motivation was somewhat weakened in, e.g., *stov-ykl-a* 'camp' ← *stov-ė-ti* 'stand'.[36]

---

[34] The stem *mezg-* is used as a base: inf. *megz-ti*, prs. 3 *mezg-a*, pst. 3 *mezg-ė* 'knit'.

[35] The suffix *-y-* is found only in the infinitive stem and functions as an inflection class marker: inf. *dar-y-ti*, prs. 3 *dar-o*, pst. 3 *dar-ė*.

[36] The suffix *-ė(j)-* is found in the infinitive and the past stem and functions as an inflection class marker: inf. *stov-ė-ti*, prs. 3 *stov-i*, pst. 3 *stov-ėj-o*.

6.2. Productivity measures of the place noun suffixes are presented in Table 7.

Table 7. **Productivity measures of deverbal place noun suffixes**

Suffix	Types	Hapaxes	Total frequency	Potential productivity (*10³)
–ykl-a	189	41	1,402,488	0.0292
–tuv-ė	33	7	458,390	0.0153

Just as indicated in the grammars, *–ykl-a* is clearly dominant and the significantly lower productivity of *-tuv-ė* can be explained by the fact that this suffix appears to be more specialized in deriving instrument nouns. In our sample, there are 123 types of instruments in *-tuv-ė* and only 33 types of place nouns with the same suffix. In the case of *–ykl-a*, however, perhaps just a handful of examples may be interpreted as possible instruments, consider *kab-ykl-a* 'clothes rack, stand' ← *kab-in-ti* 'hang';[37] it should be also noted that *–ykl-a* is not listed as an instrument suffix in the main grammars.

The Database of Lithuanian Neologisms contains only a few formations in *–ykl-a* and *-tuv-ė* (5 and 2, respectively) and their productivity is evaluated as limited (A l e k s a i t ė 2022, 81–85). As seen from the hapax counts in our corpus, especially *–ykl-a* should be interpreted as having evident derivational capacity, consider hapaxes *krov-ykl-a* 'a place to charge electric vehicle' ← *krau-ti* 'charge',[38] *išpardav-ykl-a* 'shop for selling-out (clothes)' ← *išparduo-ti* 'sell out',[39] *supirk-ykl-a* 'a place for buying-up smth.' ← *supirk-ti* 'buy up'.

7. Celebration nouns
7.1. Lithuanian word-formation descriptions traditionally distinguish a category of derived nouns designating celebrations, ceremonies, and other significant events, e.g., *ves-tuv-ės* 'wedding' ← *ves-ti* 'marry (about a man)',

---

[37] The suffix *-in-* of the base is omitted in the derivative.

[38] The past stem *krov-* is used as a base: inf. *krau-ti*, prs. 3 *kraun-a*, pst. 3 *krov-ė*.

[39] The past stem *išpardav-* is used as a base: inf. *išparduo-ti*, prs. 3 *išparduod-a*, pst. 3 *išpardav-ė*.

laido-tuv-ės 'funeral' ← *laido-ti* 'bury', *įkur-tuv-ės* 'housewarming' ← *į-si-kur-ti* 'moving to a new home, location'[40] (U r b u t i s 1965, 421; 2006, 144; S t u n d ž i a 2016, 3097). The category is generally not productive, but we decided to take a look at the most salient suffix *-tuv-ės* found in all the examples above and in some neologisms (A l e k s a i t ė 2022, 96). This suffix derives deverbal plurale tantum nouns in the nom. pl. *-ės* and uses both suffixal and non-suffixal verbal bases as illustrated above: *ves-ti*, *įsikur-ti* are non-suffixal, while *laid-o-ti* bears a suffix. The grammars note that there are about 50 more or less known formations in *-tuv-ės* in standard Lithuanian (U r b u t i s 1965, 422; 2006, 144).

Nouns with *-tuv-ės* are usually derivationally transparent. As for their semantics, there is a continuum ranging from established celebrations and ceremonies, like 'wedding' and 'funeral' above, to all kinds of events that are seen as significant, sometimes ad hoc. Some of the formations also have an ironic use, consider *(mero posto)* **daly-tuv-ės** 'sharing (of the post of the mayor)' (← *daly-ti-s* 'share'; subcorpus of Lithuanian internet texts) used jokingly in the context of parties wrangling for the mayor's post after the elections. Sometimes one may also contrast the formations in *-tuv-ės* with the unmarked action nominals in *-im-as/-ym-as* to note the difference between the noun referring to a significant event vs. the neutral action nominal, e.g., *skers-tuv-ės* 'slaughtering (of a pig)' is marked (denoting a significant event that involves a certain feasting) in contrast to the neutral *skerd-im-as* 'slaughtering' ← *skers-ti* 'slaughter'.[41]

7.2. Productivity data of *-tuv-ės* formations is presented in Table 8. We can only compare the productivity measures with other categories and conclude that *-tuv-ės*, with respect to realized and expanding productivity, is, e.g., similar to the instrument nouns in *-tuk-as* and *-ykl-ė* (116 types and 23 hapaxes and 102 types and 21 hapaxes, respectively) or to the action nominals of lower productivity in *-es-ys*, *-yb-a*, and *-ul-ys* (154, 116, 89 types and 27, 29, 23 hapaxes, respectively).

---

[40] The reflexive affix *-si-* is not transferred to the derivative.

[41] The past stem *skerd-* is used as a base for *skerd-im-as* (inf. *skers-ti*, prs. 3 *skerdži-a*, pst. 3 *skerd-ė*).

Table 8. **Productivity measures of the deverbal celebration noun suffix** *-tuv-ės*

Suffix	Types	Hapaxes	Total frequency	Potential productivity (*10^3)
-tuv-ės	99	24	151,896	0.1580

8. Conclusion

Our corpus study shows that the ordering of productive Lithuanian deverbal noun suffixes according to their realized and expanding productivity mostly corresponds to the rankings suggested in the grammars. The suffixes were also evaluated according to potential productivity, but we were cautious about interpreting the results due to significant differences in the total frequencies of the derivatives. The total frequencies appear to distort the rankings and a further study of potential productivity estimated at the same total frequencies is needed.

For action nominals, the grammars lacked approximate numbers of formations with many suffixes, but their ranking coincides with the one established in our study, in respect to both realized and expanding productivity. The grammars note that numerous formations in *-yb-a* are neologisms, but we suggest that this was the perception of the 1960s because only a small portion of formations in our corpus can be considered new nowadays. Similar to other studies, we also observed that a number of formations with the borrowed suffix *-acij-a* are real derivatives and not just borrowings that are interpretable as derived.

In the case of agent nouns, we proposed a new approach where the productivity of masculine and feminine formations is measured separately, which allowed us to note a higher realized and expanding productivity of masculine formations. This divergence is at least partly explainable by the generic use of masculine formations but further research is needed. Our lemmatization of corpus data was limited by the problem of morphological ambiguity of the homographic forms of masculine and feminine nouns, especially in the case of formations in *-toj-* and *-ėj-*. The type counts currently remain imprecise, but we manually corrected the hapax counts, so the ranking according to expanding productivity is quite reliable. Notably, the latter ranking corresponds to the one proposed in the grammars. We also demonstrated that our corpus data show a bigger difference between the type counts of *-ėj-*

and *–ik–* formations (suffix-*ėj-* is much more productive) than suggested in the grammars, which apparently base their judgements on dictionary data. When analyzing agent formations, we attempted to separate instrument nouns formed with the prototypical agent suffixes *–toj–as* and *–ėj–as*. Our data suggest that their formation is not as productive as that of the regular most productive instrument suffixes *–tuv–as* and *–ikl–is*, but can still be considered quite significant. As for other instrument nouns, our study allowed us to rank the formations with the borrowed suffix *–ator–ius*: they are third according to realized productivity, but only fifth according to expanding productivity, demonstrating their low portion of new formations. The ranking of other instrument suffixes corresponds to the one suggested in the grammars.

The result and object nouns were interesting with respect to formations in *–al–as*, which rank second according to realized productivity after formations in *–in–ys*, just as estimated by the grammars, but first according to expanding productivity. The latter productivity effect might be due to their pejorative use: it seems that speakers tend to coin new formations with this suffix to convey their negative attitude towards the results (objects) of the actions. We also measured the productivity of one borrowed suffix, namely *–at–as*, which ranked third after two of the abovementioned native suffixes.

The study of deverbal place nouns confirmed the higher realized and expanding productivity of *–ykl–a* vs. *–tuv–ė*. The corpus data also showed a certain liveliness of the category of deverbal nouns denoting celebrations and other significant events in *–tuv–ės*.

PRIESAGINIŲ VEIKSMAŽODINIŲ DAIKTAVARDŽIŲ PRODUKTYVUMAS JUNGTINIO LIETUVIŲ KALBOS TEKSTYNO DUOMENIMIS

Santrauka

Paskutiniame XX a. dešimtmetyje buvo pradėta plėtoti darybinio produktyvumo matavimo metodika, grįsta tekstynų duomenimis (B a a y e n 1992; 1993; 2009). Šiame straipsnyje imamasi tą metodiką taikyti nagrinėjant produktyviausius lietuvių kalbos priesaginių veiksmažodinių daiktavardžių darybos tipus Jungtiniame lietuvių kalbos tekstyne, sudarytame iš 3 patekstynių (iš viso 1,3 mlrd. žodžių pavartojimų), žr. D a d u r k e v i č i u s , P e t r a u s k a i t ė (2020, 123–124). Tyrimui atrenkant darybos ti-

pus remtasi pagrindinėmis lietuvių kalbos gramatikomis (U r b u t i s 1965; 2006), o aptariant gautus rezultatus atsižvelgta ir į pastarųjų metų naujadarų analizės duomenis (V a s k e l i e n ė 2017; M u r m u l a i t y t ė 2021; A l e k s a i t ė 2022). Pasitelkus minėtąją metodiką vertinamas realizuotasis, plėtros ir potencinis produktyvumas.

Atliktas tyrimas rodo, kad nagrinėtų daiktavardžių priesagų rikiuotė pagal realizuotąjį ir plėtros produktyvumą dažniausiai atitinka gramatikose (ir daugelyje kitų šaltinių) siūlomą eiliškumą. Į potencinio produktyvumo rezultatus žiūrėjome atsargiai, mat, kaip yra pastebėję ir kiti tyrėjai, dideli vedinių bendrojo dažnio skirtumai kai kuriais atvejais iškreipia produktyvumo reitingą. Manytume, kad ateityje reikėtų atlikti papildomą tyrimą, kuriame potencinis produktyvumas būtų vertinamas parenkant vienodas bendrųjų dažnių vertes, kaip tai daroma, pvz., studijose Š t i c h a u e r (2009), V a r v a r a (2019).

Veiksmų pavadinimų kategorijai mūsų tyrimas naudingas pirmiausia kiekybiniais duomenimis – dalies priesagų (–im-as, –ym-as, –yb-a, –ul-ys) apytikrio vedinių skaičiaus gramatikose nebuvo pateikiama. Taip pat atkreipėme dėmesį į tai, kad daugelio priesagos –yb-a vedinių naujadarais, kaip teigiama gramatikose, laikyti jau nebegalima – nuo tada, kai buvo parengtas „Lietuvių kalbos gramatikos“ daiktavardžių darybos skyrius (U r b u t i s 1965), prabėgo nemažai laiko, daugelis vedinių tapo visai įprasti. Su priesaga –yb-a naujai išsivedamų žodžių vis dar pasitaiko, bet apskritai šio afikso plėtros produktyvumas panašus į –es-ys ir –ul-ys. Mūsų duomenys taip pat patvirtino ir naujadarų tyrimų įžvalgas, kad priesaga –acij-a pasitaiko ne tik skoliniuose, bet ir kai kuriuose pačiõs lietuvių kalbos dirvoje pasidarytuose vediniuose.

 Tirdami veikėjų pavadinimus pasiūlėme naujovę – vyriškosios ir moteriškosios giminės vedinių produktyvumą įvertinome atskirai. Tai leido pastebėti akivaizdžiai didesnį vyriškosios giminės daiktavardžių plėtros produktyvumą – spėjame, kad tokį skirtumą bent iš dalies galėjo nulemti apibendrintoji vyriškosios giminės vedinių vartosena, bet šiai įžvalgai pagrįsti reikėtų detalesnių tyrimų. Nagrinėdami priesagų –toj-as ir –ėj-as vedinius pabandėme atskirti įrankio reikšmę turinčius daiktavardžius, kurie bendrinėje kalboje laikomi nenorminiais. Mūsų tekstyno duomenys rodo, kad įrankių pavadinimų vedyba su šiomis priesagomis nėra tokia produktyvi kaip su –tuv-as ar –ikl-is, bet apskritai vertintina kaip gana gyva ir tyrinėtina plačiau atsižvelgiant į šaltinių pobūdį, taip pat redaguotos ir neredaguotos kalbos skirtį.

Rezultatų ir objektų pavadinimų analizės rezultatai pasirodė įdomūs tuo, kad priesagų –in-ys ir –al-as eilė skyrėsi pagal produktyvumo tipus: pagal realizuotąjį pirmoji buvo –in-ys, o pagal plėtros – –al-as. Gali būti, kad didesnį plėtros produktyvumą nulėmė pejoratyvinis priesagos –al-as atspalvis, kurio kalbėtojams, regis, dažniau prisireikia spontaniškai sudarant objektų ir rezultatų pavadinimus.

Veiksmažodinių vietos pavadinimų tyrimas patvirtino akivaizdžiai didesnį priesagos –ykl-a realizuotąjį ir plėtros produktyvumą palyginti su –tuv-ė. Mūsų tekstyno duomenys

taip pat rodo, kad priesaga *-tuv-ės* gana gyvai vartojama sudarant daiktavardžius, žymin-čius įvairias šventes ar kitus reikšmingus įvykius, ir savo produktyvumu, pavyzdžiui, primena įrankių pavadinimų priesagas *-tuk-as* ir *-ykl-ė*.

ABBREVIATIONS

3 – 3rd person

ACC – accusative

F – feminine

GEN – genitive

INF – infinitive

ITR – intransitive

M – masculine

NOM – nominative

PL – plural

PRS – present

PST – past

SG – singular

TR – transitive

SOURCES

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).

BIBLIOGRAPHY

Aleksaitė, Agnė 2022, *Lietuvių kalbos naujažodžių daryba (2011–2019 m. Naujažodžių duomenyno pagrindu)*, humanitarinių mokslų daktaro disertacija, Vilnius: Lietuvių kalbos institutas.

Arkadiev, Peter 2010, Notes on the Lithuanian restrictive, *Baltic Linguistics* 1, 9–49 (https://doi:10.32798/bl.434).

Arkadiev, Peter 2011, On the aspectual uses of the prefix *be-* in Lithuanian, *Baltic Linguistics* 2, 37–78 (https://doi:10.32798/bl.426).

Arkadiev, Peter 2013, On discontinuative action nominals in Lithuanian: Addendum et corrigendum to Arkadiev (2011), *Baltic Linguistics* 4, 215–218 (https://doi.org/10.32798/bl.414).

Baayen, Harald R. 1992, Quantitative aspects of morphological productivity, in Geert Booij, Jaap van Marle (eds.), *Yearbook of Morphology 1991*, Dordrecht: Kluwer, 109–149 (https://doi.org/10.1007/978-94-011-2516-1_8).

Baayen, Harald R. 1993, On frequency, transparency and productivity, in Geert Booij, Jaap van Marle (eds.), *Yearbook of Morphology 1992*, Dordrecht: Kluwer, 181–208 (https://doi.org/10.1007/978-94-017-3710-4_7).

Baayen, Harald R. 2009, Corpus linguistics in morphology: Morphological productivity, in Anke Lüdeling, Merja Kytö (eds.), *Corpus Linguistics: An International Handbook* 2, Berlin, New York: Mouton de Gruyter, 899–919 (https://doi.org/10.1515/9783110213881.2.899).

Dadurkevičius, Virginijus 2017, Lietuvių kalbos morfologija atvirojo kodo „Hunspell" platformoje, *Bendrinė kalba* 90, 1–17 (https://journals.lki.lt/bendrinekalba/article/view/156).

Dadurkevičius, Virginijus, Rūta Petrauskaitė 2020, Corpus-based methods for assessment of traditional dictionaries, in Andrius Utka, Jurgita Vaičenonienė, Jolanta Kovalevskaitė, Danguolė Kalinauskaitė (eds.), *Human Language Technologies – The Baltic Perspective*, Amsterdam: IOS Press, 123–126 (https://doi.org/10.3233/FAIA200613).

Dal, Georgette, Fiammetta Namer 2016, Productivity, in Andrew Hippisley, Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, Cambridge: Cambridge University Press, 70–90 (https://doi.org/10.1017/9781139814720.004).

Evert, Stefan, Marco Baroni 2007, zipfR: Word Frequency Modeling in R, in Sophia Ananiadou (ed.), *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Demo and Poster Sessions*, Prague: Association for Computational Linguistics, 29–32 (https://aclanthology.org/P07-2008).

Gaeta, Livio 2015, Action nouns in Romance, in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe* 2, Berlin, Boston: De Gruyter Mouton, 1209–1229 (https://doi.org/10.1515/9783110246278-024).

Gaeta, Livio, Davide Ricca 2006, Productivity in Italian word formation: a variable-corpus approach, *Linguistics* 44(1), 57–89 (https://doi.org/10.1515/LING.2006.003).

Gaeta, Livio, Davide Ricca 2015, Productivity, in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe* 2, Berlin, Boston: De Gruyter Mouton, 842–858 (https://doi.org/10.1515/9783110246278-003).

Inčiuraitė-Noreikienė, Lina 2015, Lietuvių kalbos dūriniai su neoklasikiniais dėmenimis, *Baltistica* 50(2), 245–259 (https://doi.org/10.15388/baltistica.50.2.2238).

Inčiuraitė-Noreikienė, Lina 2017, *Dabartinės lietuvių kalbos vardažodžių su (neo) klasikiniais elementais morfeminė sandara ir daryba*, humanitarinių mokslų daktaro disertacija, Vilnius: Vilniaus universitetas.

Jašinskaitė, Irena 1971, Veiksmažodžių struktūriniai tipai ir daryba, in Kazys Ulvydas (ed.), *Lietuvių kalbos gramatika* 2, Vilnius: Mintis, 218–269.

Kietytė, Samanta 2023, Productivity of prefixation: A case of Lithuanian prefixes *nu-*, *pri-*, *į-* and *iš-*, in Andra Kalnača (ed.), *Valoda: nozīme un forma* 14, Rīga: Latvijas Universitātes Akadēmiskais apgāds, 91–110 (https://doi.org/10.22364/vnf.14).

Koptjevskaja-Tamm, Maria 2015, Action nouns, in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe* 2, Berlin, Boston: De Gruyter Mouton, 1195–1209 (https://doi.org/10.1515/9783110246278-023).

Melloni, Chiara 2015, Result nouns, in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe* 2, Berlin, Boston: De Gruyter Mouton, 1253–1268 (https://doi.org/10.1515/9783110246278-027).

Mikelionienė, Jurgita 2000, *Naujoji lietuvių kalbos leksika (1991–1996 m. kompiuterinio periodikos tekstyno pagrindu)*, humanitarinių mokslų daktaro disertacija, Kaunas: Vytauto Didžiojo universitetas.

Miliūnaitė, Rita, Agnė Aleksaitė 2011, *Lietuvių kalbos naujažodžių duomenynas*, Vilnius: Lietuvių kalbos institutas (http://naujazodziai.lki.lt).

Mühleisen, Susanne 2015, Patient nouns, in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe* 2, Berlin, Boston: De Gruyter Mouton, 1316–1327 (https://doi.org/10.1515/9783110246278-031).

Murmulaitytė, Daiva 2014, Priesagos *-iklis* naujadarų apibrėžčių sistemiškumas, *Bendrinė kalba* 87, 1–17.

Murmulaitytė, Daiva 2016, Naujieji asmenų pavadinimai darybos ir semantiniu aspektu, *Lietuvių kalba* 10, 1–22 (https://doi.org/10.15388/LK.2016.22591).

Murmulaitytė, Daiva 2021, *Naujažodžių darybos ir morfemikos tyrimų perspektyvos (Lietuvių kalbos naujažodžių duomenyno atvejis)*, Vilnius: Lietuvių kalbos institutas (https://doi.org/10.35321/e-pub.16.naujadaros-tyrimu-perspektyvos).

Pakerys, Jurgis, Virginijus Dadurkevičius, Agnė Navickaitė-Klišauskienė 2024, How lemmatization and derivational annotation affect productivity measures in corpus: The case of Lithuanian deverbal agent nouns, in Andra Kalnača (ed.), *Valoda: nozīme un forma* 15, Rīga: Latvijas Universitātes Akadēmiskais apgāds, 138-151 (https://doi.org/10.22364/vnf.15.09).

Paulauskienė, Aldona, Vitas Labutis, Aleksandras Vanagas, Jonas Žemaitis, Jonas Šukys 1976, *Kalbos praktikos patarimai*, Vilnius: Mokslas.

Pupkis, Aldonas 1980, *Kalbos kultūros pagrindai*, Vilnius: Mokslas.

Rainer, Franz 2015, Agent and instrument nouns, in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe* 2, Berlin, Boston: De Gruyter Mouton, 1304–1316 (https://doi.org/10.1515/9783110246278-030).

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 Fabio Montermini, Gilles Boyé, Jesse Tseng (eds.), *Selected Proceedings of the 6th Décembrettes: morphology in Bordeaux*, Somerville, MA: Cascadilla Proceedings Project, 138–147.

Stundžia, Bonifacas 2016, Lithuanian, in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe* 5, Berlin, Boston: De Gruyter Mouton, 3089–3106 (https://doi.org/10.1515/9783110424942-001).

Szymanek, Bogdan 2015, Place nouns, in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe* 2, Berlin, Boston: De Gruyter Mouton, 1327–1339 (https://doi.org/10.1515/9783110246278-032).

Urbutis, Vincas 1965, Daiktavardžių daryba, in Kazys Ulvydas (ed.), *Lietuvių kalbos gramatika* 1, Vilnius: Mintis, 251–473.

Urbutis, Vincas 1978, *Žodžių darybos teorija*, Vilnius: Mokslas.

Urbutis, Vincas 2006, Daiktavardžių daryba, in Vytautas Ambrazas (ed.), *Dabartinės lietuvių kalbos gramatika*, Vilnius: Mokslo ir enciklopedijų leidybos institutas, 86–166.

Van Marle, Jaap 1992, The relationship between morphological productivity and frequency: A comment on Baayen's performance-oriented conception of morphological productivity, in Geert Booij, Jaap van Marle (eds.), *Yearbook of Morphology 1991*, Dordrecht: Springer, 151–163 (https://doi.org/10.1007/978-94-011-2516-1_9).

Varvara, Rossella 2019, Misurare la produttività morfologica: i nomi d'azione nell'italiano del ventunesimo secolo, in Bruno Moretti, Aline Kunz, Silvia Natale, Etna Krakenberger (eds.), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Socie̜ta di Linguistica Italiana* (Berna, 6–8 settembre 2018), Milano: Officinaventuno, 187–201 (https://doi.org/10.17469/O2102SLI000011).

Vaskelienė, Jolanta 2003, Sigito Gedos individualūs dariniai („Sokratas kalbasi su vėju"), *Žmogus ir žodis* 1, 43–48.

Vaskelienė, Jolanta 2007, Žodynų nefiksuoti dariniai ir jų funkcijos Mariaus Katiliškio novelių rinkinyje „Seno kareivio sugrįžimas", *Valoda – 2007. Valoda dažādu kultūru kontekstā*, Daugavpils: Saule, 505–512.

Vaskelienė, Jolanta 2011, Žodynų nefiksuotų darinių funkcionavimas Jono Strielkūno lyrikoje, *Res humanitariae* 10, 54–70.

Vaskelienė, Jolanta 2012, Žodynų nefiksuotų darinių funkcionavimas Ričardo Gavelio kūriniuose, *Filologija* 17, 145–162.

Vaskelienė, Jolanta 2017, Lietuvių rašytojų naujadarų darybos ir semantikos ypatumai, *Bendrinė kalba* 90, 1–30.

Zeldes, Amir 2012, *Productivity in Argument Selection: From Morphology to Syntax*, Berlin, Boston: De Gruyter Mouton (https://doi.org/10.1515/9783110303919).

Jurgis PAKERYS, Agnė NAVICKAITĖ-KLIŠAUSKIENĖ
Baltistikos katedra
Vilniaus universitetas
Universiteto g. 5
LT-01513 Vilnius
Lithuania
[*jurgis.pakerys@flf.vu.lt*]
[*agne.navickaite@flf.vu.lt*]

Virginijus DADURKEVIČIUS
Skaitmeninių išteklių ir tarpdisciplininių tyrimų institutas
Vytauto Didžiojo universitetas
V. Putvinskio g. 23
LT-44243 Kaunas
Lithuania
[*virginijus.dadurkevicius@vdu.lt*]