

USAGE AND EMPIRICAL PRODUCTIVITY OF INTERNATIONAL ADJECTIVAL SUFFIXES IN SLOVAK BASED ON GENERAL AND SPECIALISED CORPORA

JANA LEVICKÁ

Slovak National Corpus, E. Štúr Institute of Linguistics, Slovak Academy of Sciences, Bratislava, Slovakia

LEVICKÁ, Jana: Usage and empirical productivity of international adjectival suffixes in Slovak based on general and specialised corpora. *Journal of Linguistics*, 2021, Vol. 72, No 2, pp. 394 – 404.

Abstract: The paper attempts to identify the usage and productivity of five different international suffixes in Slovak by means of corpus evidence. The analysis focuses on real and potential productivity in a two-stage comparison: 1) tokens/lemmas occurring in a general balanced corpus vs general corpus of specialised and academic texts, 2) general corpus of specialised and academic texts vs specialised (sub)corpora of medical, legal, economic and religious texts. The aim of the analysis is to explore whether productivity varies across registers by means of statistical measures.

Keywords: productivity, realized productivity, potential productivity, general corpus, specialised corpus, adjective, suffix

1 INTRODUCTION

In the last two decades, terminology research has seen the emergence of a new research topic: variation analysis. One of the most variant-productive areas is that resulting from the clash of two contending tendencies: internationalisation and naturalisation, i.e., coining of new terms from national language resources. In Slovak, usage of the so-called international loanwords¹ has deep historical roots, as Latin was the official language of the Hungarian Kingdom, territory which until 1867 also included Slovakia. A high proportion of words from Latin can be found also in the general Slovak lexicon [1, p. 81], therefore, many Slovaks find borrowing and usage of international words and terms (especially via English) natural.

However, there has always been a natural tendency to coin Slovak counterparts to international words, the realm of terminology including. The clash between these internationalising and naturalising tendencies often results in competing or coexisting (synonymous) words and terms [2, p. 273]. It has been observed that this nationalising tendency is not uniformly present in specialised domains [3, p. 174]; its influence

¹ Terms of Latin and/or Greek origin, occurring at least in three genetically unrelated languages, which are more or less adapted to Slovak [4, p. 89]. More information on corpora can be found in part 2.

manifests itself unevenly, depending on the tradition and character of individual disciplines or domains, as well as on their linguistic and word-forming specificities.

Furthermore, in addition to the traditional preoccupation with nouns, terminology research has also shifted its attention to other parts of speech. It was during the 1990s and after 2000 that adjectives began to emerge into the limelight. Their analyses, underlying differentiating and classifying functions, as well as their ability to *build* multi-word terms, started to be published. The significance of adjectives in terminology and language for specialised purposes (LSP) can be supported also by corpus evidence. While Czech data indicate the presence of as much as 13% of adjectives in LSP texts of SYN 2000 compared to other parts of speech [5], the ratio of adjectives in the majority of specialised corpora of the Slovak National Corpus project amounts only to no more than 9%. However, as can be seen in the table 1, this ratio is at least 2% higher when compared to the reference corpus. Only religious corpus blf-2.0 features roughly the same percentage distribution of adjectives as the reference corpus (7.46%). The third row of table 1 presents the frequency and ratio of gerunds that are also used with differentiating and classifying functions in multi-word terms.

corpus	prim-7.0-frk	prim-9.0-public-prf	prim-9.0-juls-all (MED)	legal-1.1.ver.bz 1991_2011	blf-2.0	ecn-2.0-public
tokens/corpus	253,137,609	149,581,785	7,099,555	33,600,183	65,920,357	164,987,015
tokens/adjectives	19,090,396 7.54%	13,415,554 8.97%	660,025 9.3%	4,841,400 9.88%	4,914,860 7.46%	15,540,381 9.42%
tokens/gerunds	3,174,567 1.25%	2,394,914 1.6%	112,164 1.58 %	1,267,598 2.59%	761,719 1.16%	2,154,124 1.31%

Tab. 1. Number of tokens and ratios of adjectives and gerunds in reference and specialised corpora of the SNC project

Coexisting and competing forms can also be found among adjectives in multi-word terms or in specialised discourse as such. In general, these forms include Slovak counterparts to international adjectives (e.g. *vnútrožilový* – *intravenózný* ‘intravenous’), the latter group comprises also a subgroup that shows the same two tendencies by means of variation of international and Slovak affixes combined with the same international roots (e.g., *bakteriálny* – *baktériový* ‘bacterial’).

2 RESEARCH AIMS AND DATA

This paper is an attempt to identify the usage and productivity of five different international suffixes in Slovak by means of corpus evidence. Moreover, the analysis

will focus on a two-stage comparison: 1) tokens/lemmas occurring in a general balanced corpus vs general corpus of specialised and academic texts, 2) general corpus of specialised and academic texts vs specialised (sub)corpora of medical, legal, economic, and religious texts. These corpora may help to explore whether productivity varies across registers.

Suffixes selected for analysis are representative of a minor group within a range of adjectival suffixes.² In terms of their composition, the five suffixes may be termed ‘reduplicated’ as they consist of an (adapted) international adjectival suffix combined with semantically equivalent Slovak **-ný/ny**: **-álny**, **-árny**, **-itný**, **-ívny**, **-ózny**. Obviously, all of them come from Latin suffixes used to coin adjectives from Latin nouns (**-alis** with the variant **-aris**, **-ītus**, **-ōsus**) or verbs (**-ōrius**). It is worth pointing out, however, that many Slovak adjectives with analysed suffixes entered the Slovak lexicon via French and English.

2.1 Corpora used in the analysis

All three corpora and three subcorpora used in this analysis were released by the Department of the SNC in 2013–2020 and are accessible for all registered users.

The first one, the reference corpus *prim-7.0-frk* [7], amounting to more than 253 million tokens, is composed of an even share of journalistic, specialised, and fictional texts (64.12% of them are Slovak while 29.51% represent translations) written from 1991–2015. The corpus was used in the compilation of two frequency dictionaries of Slovak (2017, 2018), as well as the reverse dictionary (2018).

The second corpus, *prim-9.0-public-prf* [8], is a publicly available subcorpus of the primary corpus of the SNC project. Compiled from specialised, academic, and non-fiction texts, this subcorpus features more than 149 million tokens, and documents general discourse of science and research, including specialised journalism. Its texts were written between 1955 and 2019.

The smallest subcorpus of all searched corpora exists as a result of filtering the primary corpus of the SNC project [9], version 9.0. It consists of texts that belong to the field of medicine, written from 1976 to 2019, and comprises slightly more than 7 million tokens.

In order to create a comparable source with other specialised corpora and the reference corpus, the specialised corpus *legal-1.1* [10], built in cooperation with the Slovak Ministry of Justice, was narrowed down to legislative texts created between 1991 and 2011. Its approximately 49 million tokens were thus reduced to 33.5 million tokens.

The specialised corpus – *blf-2.0* [11] – focusing on the field of religion – was released in 2014. Its texts consist of almost 66 million tokens written from 1989–

² Ološtiak and Ološtiaková [6, p. 230] mention as many as 38 suffixes, though 25 suffixes in their sample – derived from the *Slovník koreňových morfém slovenčiny* comprising 66,500 analysed lexical units – represent only 1% of adjectives.

2014. It comprises more than 80% of thematic journals and newspapers. Similarly, specialised corpus ecn-2.0-public [12], devoted to the field of economics, includes as much as 96.24% of specialised texts published in thematic journals and newspapers. Texts of this corpus come from 1992–2014 and comprise almost 165 million tokens.

3 THEORETICAL BACKGROUND AND METHODOLOGY

Morphological productivity represents one of the most contentious linguistic issues and is the focus of extensive research and discussions.³ One complex theory of word-formation productivity was presented by a Czech linguist, Miloš Dokulil, in his work *Teorie tvoření slov* in 1962 [13], which also included the differentiation of the *systemic* and the *empirical productivity* (also termed ‘parole’ or ‘real productivity’ [14]), the latter determined by extra-linguistic factors. With the availability of extensive corpora, it is possible to measure, identify, and analyse the concept of Dokulil’s empirical productivity of a word-formation type or element in a language at a given time. Dokulil believed that even “approximative data concerning the quantitative use of a given word-formation process or element are of paramount importance in the overall picture of a given language in general and for the characteristics of its lexicon in particular” [Ibid., p. 77]. Moreover, Dokulil’s theory is also inspirational in the differentiation of the so-called *absolute frequency* of word-formation processes, types, and elements and the *relative frequency*, which is register- or domain-dependent.

Echoing Dokulil’s theory of productivity, some contemporary Czech linguists elaborate on and verify his assumptions on corpus data (see, e.g., [14], [15], [16], [17] or [18]). Štícha advocates the need to analyse suffixes one by one, to identify the frequency and ratio of derived words, coined with these suffixes, in a given corpus [14, p. 100]. Štícha proposes the analysis of empirical/parole productivity not only on big corpora, but also on a series of corpora of different size and composition [Ibid., p. 104]. Corpus findings and statistical data indicate that empirical/parole productivity could be differentiated further into general and specialised categories which will be of special interest in the context of this study [Ibid.].

The main “trend” concerning statistical-based research of productivity, introduced especially by the studies of Harald Baayen and his colleagues in the 1990s ([19], [20], [21], [22], [23]), claims an importance for *hapax legomena* in a given corpus in determining the degree of productivity of a word-formation type or element. The rationale behind this method is that lemmas which occur only once in

³ More information can be found, e.g., in Hulse, V.: Productivity in morphological negation: a corpus based approach. The University of Manchester (2011). Available at: [https://www.research.manchester.ac.uk/portal/en/theses/productivity-in-morphological-negation-a-corpusbased-approach\(266d2241-a266-4b99-8fab-e19571381d8f\).html](https://www.research.manchester.ac.uk/portal/en/theses/productivity-in-morphological-negation-a-corpusbased-approach(266d2241-a266-4b99-8fab-e19571381d8f).html).

a corpus are indicative of the creation of new words. However, many researchers criticize the significance attributed to hapaxes in measuring morphological productivity and emphasize the fact that not all hapaxes represent new coinages, on the contrary, this group often consists of peripheral lexical units including archaisms or words that simply happen to occur only once in a given corpus. To do Baayen justice, he explicitly states that hapaxes only correlate to the number of neologisms and that “they only function as a tool for a statistical *estimation method* aimed at gauging the rate of expansion of morphological categories” [24, p. 906].

In his 2009 paper [24], Baayen proposed a more elaborate theory of (morphological) productivity distinguishing three levels:

1) realized productivity, which reflects the productivity of a word-formation type/process in the past and which can be “estimated by the type count”, i.e., the number of lemmas in a corpus;

2) expanding productivity, which is an estimate of the contribution of a morphological category to the growth rate of the total vocabulary. Baayen suggests it be calculated as the ratio of hapaxes with affix X/all hapaxes in a corpus;

3) potential productivity, which enables one to “estimate the growth rate of the vocabulary of the morphological category itself” and can be calculated as the number of hapaxes with affix X/tokens with affix X. Baayen used this method already in [19], stressing that it is to show only the statistical probability ratio of future coinages.

Due to legitimate criticism of hapax significance mentioned earlier, this analysis is based on manually *cleaned-up* data. Lists of hapaxes extracted from every corpus and subcorpus were checked and several groups of lemmas excluded:

a) lemmas of Slovak origin including incidentally the same sequence of characters as the analysed suffixes (exclusion not only from the list of hapaxes, but also from the list of lemmas with the suffix X);

b) lemmas with typos or orthographical mistakes;

c) lemmas found in general Slovak dictionaries and the Dictionary of Foreign Words (most of them representing terms of specialised domains);

d) lemmas found in two most extensive SNC corpora: exclusion of lemmas with 3 and more occurrences in general corpus prim-9.0-juls-all [9] and in legal corpus legal-1.1 [10], provided that those occurrences come from 3 different sources and 3 different years.

Overall, the exclusion ranged from 26% up to 100% of hapaxes in individual (sub)corpora. However, the cleaned-up lists of hapaxes may still comprise lemmas that are not neologisms, due to the lack of up-to-date specialized dictionaries and the extent and content of the corpora used.

Moreover, if a list of lemmas with suffix X comprised two lemmas differing only in the usage or non-usage of a hyphen, these were merged, as well as lemmas (not being proper names) with a capitalised and non-capitalised first letter.

For practical reasons, a thorough manual check and clean-up of lists of all hapaxes from big corpora is more than time-consuming and unfeasible. Therefore, further analyses will focus only on Baayen's realized and potential productivity:

1. **realized productivity** in order to determine the productivity of suffixes with respect to past and present linguistic situations, completed with the corpus statistics reflecting their usage;

2. **potential productivity** in order to estimate the rate at which new types are to be expected to appear. However, I decided not to use the hapax/token method, but the hapax/type method in line with Van Marle's reasoning [25] that token frequency is not as relevant a variable in the measure of productivity as the number of lemmas. Moreover, I assume that an estimation of future productivity should be related to contemporary and past productivity.

As far as the corpus search is concerned, I did not base the queries on lemmatization and morphological tagging of the (sub)corpora because with Latinate words the lemmatization and tagging proved to be inadequate and erroneous. Therefore, I opted for simple search of specific ending of a token, e.g. [lemma="*.álly"], combined with automatic filtering of words with incidentally the same string of characters.

4 RESEARCH RESULTS

4.1 Usage of suffixes and realized productivity

Table 2 shows the raw frequency of tokens with analysed suffixes occurring in 6 selected (sub)corpora. For the sake of comparison, the second column features the normalised frequency (ipm) of these tokens and, thus, enables an inference of their usage in general and specialised domains. The suffixes have been listed in order of decreasing number of tokens, which is mirrored by the decreasing normalised frequency, except for the order of the pair *-ózny* and *-órny* in legal corpus and *-árny* and *-itný* in the economic corpus. Suffix *-álly* is clearly the most widely used in all (sub)corpora, while *-órny* is at the opposite end of the frequency axis in 5 (sub)corpora. Four out of five suffixes clearly reach higher normalised frequencies in the general corpus of specialised and academic texts (prim-9.0-public-prf) compared to reference corpus, as expected. Frequency differences between reference corpus and prim-9.0-public-prf, as well as those between prim-9.0-public-prf and each specialised (sub)corpus, were subjected to a test of significance test (log likelihood test), which confirmed the significance of observed statistical data with the exception of the suffix *-árny* in medical texts and the suffix *-órny* in legal texts. It is also noteworthy that normalised frequencies of suffixes in medical subcorpus equal or considerably exceed the ipm in prim-9.0-public-prf while the ipm of suffixes in religious corpus is manifestly lower than in the reference corpus.

suffix	prim-7.0-frk	prim-9.0- public-prf	prim-9.0- juls-all (MED)	legal-1.1.ver. bz 1991_2011	blf-2.0	ecn-2.0- public
	Tokens	Tokens	Tokens	Tokens	Tokens	Tokens
	IPM	IPM	IPM	IPM	IPM	IPM
-álny	547,742	537,682	31,408	102,863	128,472	597 005
	2163.81	3594.57	4423.94	3061.38	1948.9	3618.5
-árny	84,780	94,586	4,529	25,980	14,253	38,178
	334.92	634.14	637.93	773.21	216.22	231.40
-itný	44,655	30,402	2,683	2,256	9,176	52,574
	176.41	203.25	377.91	67.14	139.20	318.66
-ózny	16,531	8,863	1,416	482	2,321	9,440
	65.30	59.25	199.45	14.35	35.21	57.22
-órny	4,672	4,537	884	1,959	620	2,039
	18.46	30.33	124.51	58.30	9.41	12.36

Tab. 2. Frequency and normalised frequency of analysed suffixes in the respective (sub)corpora

The estimate of realized productivity, or the insight into the extent of past new coinages by means of analysed suffixes, can be seen in table 3, which shows not only the number of lemmas in the selected (sub)corpora, but also normalised counts of lemmas per million. Again, the first position in the table is taken by the suffix *-álny* and the last place by the suffix *-órny*. Both absolute and normalised counts of lemmas are distinctly higher in prim-9.0-public-prf compared to the reference corpus, and from among specialised (sub)corpora, it is the medical field in which all five suffixes appeared most frequently. Only the number of lemmas in economic corpus, with the exception of lemmas with *-itný*, is closer to the reference corpus than to the corpus of specialised texts. Suffixes *-itný* and *-ózny* feature very similar statistics with the exception of medical texts, in which lemmas with *-ózny* are twice as numerous per million than those with *-itný*. The same reversed order of these two suffixes, compared to other corpora, is in the reference corpus.

suffix	prim-7.0-frk		prim-9.0- -public-prf		prim-9.0- juls-all (MED)		legal-1.1.ver. bz 1991_2011		blf-2.0		ecn-2.0- public	
-álny	1,006	3.97	1,179	7.88	519	73.10	322	9.58	643	9.75	753	4.56
-árny	298	1.18	366	2.45	180	25.35	100	2.98	144	2.18	204	1.24
-itný	90	0.36	97	0.65	36	5.07	41	1.22	55	0.83	118	0.72
-ózny	115	0.45	119	0.80	72	10.14	40	1.19	52	0.79	83	0.50
-órny	25	0.09	35	0.23	13	1.83	10	0.30	20	0.30	27	0.16

Tab. 3. Number of lemmas in a given (sub)corpus followed by the same count normalised per million tokens

To complete the picture, the last table in this part presents a statistical method frequently used in analyses of (morphological) productivity – type/token ratio of analysed suffixes. These data in table 4 reflect both the existence of types with one of the suffixes, as well as the extent to which these types are used. In this two-dimensional perspective, the ranking of analysed suffixes is reversed compared to the previous table – the first place is occupied either by *-ózny* or *-órny*, while *-álny* can be found at the bottom of table in 5 (sub)corpora. However, it is not possible to compare these ratios across the corpora as they were calculated from raw token frequencies and counting of lemmas.

Type/token ratio											
prim-7.0-frk		prim-9.0-public-prf		prim-9.0-juls-all (MED)		legal-1.1.ver.bz 1991_2011		blf-2.0		ecn-2.0-public	
-ózny	0.006956627	-ózny	0.013426605	-ózny	0.050847458	-ózny	0.082987552	-órny	0.032258065	-órny	0.013241785
-órny	0.005351027	-órny	0.007714349	-árny	0.039743873	-itný	0.018173759	-ózny	0.022404136	-ózny	0.008792373
-árny	0.003514980	-árny	0.003869494	-álny	0.016524452	-órny	0.005104645	-árny	0.010103136	-árny	0.005343391
-itný	0.002015452	-itný	0.003190580	-órny	0.014705882	-árny	0.003849115	-itný	0.005993897	-itný	0.002244455
-álny	0.001836631	-álny	0.002192746	-itný	0.013417816	-álny	0.003130377	-álny	0.005004982	-álny	0.001261296

Tab. 4. Type/token ratios in a given (sub)corpus

4.2 Potential productivity

As I indicated in part 3, for calculating the potential productivity of analysed suffixes, I decided to use not the hapax/token method, but the hapax/type method, in order to emphasize the relation of hapaxes to types rather than tokens. Table 5 introduces the potential productivity ratios in decreasing order. Note first that the order of suffixes in the reference corpus and prim-9.0-public-prf overlaps only partially: while *-árny* and *-álny* keep the same 3rd and 4th place, respectively, *-órny* occupies the last place in the reference corpus, but tops prim-9.0-public-prf. Similarly, suffix *-ózny* takes the first place in the reference corpus, but the 4th place in prim-9.0-public-prf. If we compare the situation in prim-9.0-public-prf and specialised (sub)corpora, the most productive suffix seems to be *-órny*, though in medical texts, it occupies the last place of the ranking and the 3rd place in economic texts. Very different ranking can be observed for the suffix *-ózny*: 4th place in prim-9.0-public-prf and the religious corpus, 2nd place in the legal and economic corpora and 1st place in medical texts, as expected. Suffix *-itný* is either least productive (in prim-9.0-public-prf, legal and religious corpora) or the most productive (1st place in economic and 2nd place in medical texts). Relatively stable potential productivity is manifested by the suffix *-álny*: 3rd place in the ranking of three (sub)corpora, 2nd place in religious and 4th place in economic texts. The last but not least, suffix *-árny* has the 2nd highest productivity in prim-9.0-public-prf, 4th in medical and legal texts, 3rd in religious texts, but lowest in economic texts.

Hapax/type ratio											
prim-7.0-frk		prim-9.0-public-prf		prim-9.0-juls-all (MED)		legal-1.1.ver.bz 1991 2011		blf-2.0		ecn-2.0-public	
-ózný	0.252173913	-órny	0.228571429	-ózný	0.208333333	-órny	0.1	-órny	0.2	-itný	0.237288136
-árny	0.201342282	-árny	0.210382514	-itný	0.138888889	-ózný	0.075	-álly	0.188180404	-ózný	0.228915663
-álly	0.173956262	-álly	0.209499576	-álly	0.129094412	-álly	0.052795031	-árny	0.145833333	-órny	0.185185185
-itný	0.122222222	-ózný	0.201680672	-árny	0.1	-árny	0.05	-ózný	0.134615385	-álly	0.177954847
-órny	0.12	-itný	0.195876289	-órny	0.076923077	-itný	0	-itný	0.090909091	-árny	0.142156863

Tab. 5. Hapax/type ratios in a given (sub)corpus. The suffixes in each subtable are listed in order of their decreasing ratio

In order to put the data in one more perspective, let us regroup the (sub)corpora in the order of their increasing corpus size and show the ratio of hapaxes as a percentage. The aim of the reordering is to answer the question of František Štícha [14, p. 255] regarding what the ratio of hapaxes will be with the increasing size of corpora. Štícha hypothesized that a significant increase of lemmas with a specific suffix correlated with an increase of low-frequency lemmas with the same affix would testify to/indicate a high real productivity of this type in a given time. However, our (sub)corpora differ significantly not only in terms of size but also in terms of types of texts and their proportion, therefore, it is not possible to test this hypothesis fully, just to indicate discernible trends. Table 6 shows that apart from the reference corpus data, two corpora – legal and economic types – are incoherent with the increasing number of either lemmas or hapaxes, or both. We can hypothesize that the reason lies in their composition and, possibly, in the character of the discipline, e.g., legal domain is rather hesitant towards linguistic innovations. Only in the case of two suffixes, *-álly* and *-órny*, it is possible to observe both the increase of lemmas and ratio of hapaxes in at least three (sub)corpora – medical subcorpus, religious corpus and prim-9.0-public-prf. An interesting drop in data can be seen between the medical subcorpus and the legal corpus, which is almost ten times larger) – except for the number of *-itný* lemmas and the ratio of *-órny* hapaxes. Similar drop, including the exception of *-itný* lemmas and hapaxes, is between prim-9.0-public-prf data and the data from economic texts. If we narrow our focus to the difference between general and specialised texts, we can observe that the number of lemmas, as well as the hapax ratio in prim-9.0-public-prf, is higher compared to the reference corpus, except for the ratio of *-ózný* hapaxes.

suffix	prim-9.0-juls-all (MED)		legal-1.1.ver.bz 1991_2011		blf-2.0		prim-9.0-public-prf		ecn-2.0-public		prim-7.0-frk	
-álly	519	12.91%	322	5.28%	643	18.82%	1179	20.95%	753	17.8%	1006	17.4%
-árny	180	10%	100	5%	144	14.58%	366	21.04%	204	14.22%	298	20.13%
-itný	36	13.89%	41	0%	55	9.09%	97	19.59%	118	24.14%	90	12.22%
-ózný	72	20.83%	40	7.5%	52	13.46%	119	20.17%	83	22.89%	115	25.22%
-órny	13	7.69%	10	10%	20	20%	35	22.86%	27	18.52%	25	12%

Tab. 6. Ratio of hapaxes in % in (sub)corpora ordered from the smallest (medical subcorpus) to the biggest corpus (prim-7.0-frk)

To sum it up, the most productive suffix in terms of past coinages is *-álny*, especially in medical, legal and religious texts, the least productive, in this perspective, is *-órny*, though the number of different lemmas with this suffix is significantly higher in prim-9.0-public-prf and specialised (sub)corpora compared to the reference corpus.

From a future productivity point of view, the picture is less clear-cut: the most productive suffix is *-órny* but only in three corpora. In the remaining three (sub) corpora, the ranking is topped in two instances by *-ózny* and once by *-itný*. Moreover, the same suffix *-órny* seems to be least productive in future regarding general and medical texts. In as many as three (sub)corpora, it is the suffix *-itný* that has taken the final place in the productivity ranking. Similarly, the last place in economic texts productivity ranking is occupied by the suffix *-árny*.

5 CONCLUSION

A proposed analysis of word-formation productivity of selected suffixes in Slovak indicates noteworthy differences depending on domains and registers. However, these differences need to be verified in specialised corpora, balanced in terms of genres and types. An open question remains as to whether the analysis of the share of neologisms in low-frequency lemmas would not change the overall picture, as several researchers note that word-frequency distribution of productive affixes is supposed to be distinctly shifted towards low-frequency lemmas comprising new coinages.

ACKNOWLEDGEMENTS

The paper has been written within the Slovak National Corpus project supported by the Slovak Academy of Sciences, Ministry of Education, Science, Research and Sport of the Slovak Republic, Ministry of Culture of the Slovak Republic and the Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences.

References

- [1] Horecký, J. (1999). Internacionalizácia a europeizácia slovenčiny. In Internacionalizácia v súčasných slovanských jazykoch. (Ed.) J. Bosák, pages 80–82, Bratislava, Veda.
- [1] Buzássyová, K., Horecký, J., Bosák, J. et al. (1989). Dynamika slovnej zásoby súčasnej slovenčiny. Bratislava, Vydavateľstvo Slovenskej akadémie vied.
- [2] Jedlička, A. (1974). Spisovný jazyk v současné komunikaci. Praha, Univerzita Karlova.
- [3] Buzássyová, K. (1991). Opakovaná internacionalizácia a problém identifikácie morfológických a lexikálnych jednotiek. Jazykovedný časopis, 42(2), pages 89–104.
- [4] Bozděchová, I. (2009). Současná terminologie (se zaměřením na kolokační termíny z lékařství). Praha, Karolinum.

- [5] Ološtiak, M., and Ološtiaková, L. (2015). Formálno-procesuálne aspekty slovotvornej motivácie. In *Kvalitatívne a kvantitatívne aspekty tvorenia slov v slovenčine*, pages 207–308, Prešov, Filozofická fakulta Prešovskej univerzity v Prešove.
- [6] Slovenský národný korpus. Korpus prim-7.0-frk. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2018. Accessible at: <https://korpus.sk>.
- [7] Slovenský národný korpus. Korpus prim-9.0-public-prf. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2020. Accessible at: <https://korpus.sk>.
- [8] Slovenský národný korpus. Korpus prim-9.0-juls-all. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2020. Accessible at: <https://korpus.sk>.
- [9] Slovenský národný korpus. Korpus legal-1.1. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2013. Accessible at: <https://korpus.sk>.
- [10] Slovenský národný korpus. Korpus blf-2.0. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2014. Accessible at: <https://korpus.sk>.
- [11] Slovenský národný korpus. Korpus ecn-2.0-public. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2016. Accessible at: <https://korpus.sk>.
- [12] Dokulil, M. (1962). *Teorie tvoření slov*. Praha, Nakladatelství ČSAV.
- [13] Štícha, F. (2012). Jak v epoše elektronických korpusů následovat Miloše Dokulila (Miloši Dokulilovi ke stému výročí narození). In *Jazykovědné aktuality*, 49, pages 95–107.
- [14] Štícha, F. (2002). K Dokulilovu pojmu slovotvorné produktivity (z hlediska korpusové analýzy). In *Čeština doma a ve světě*, 4, pages 302–310.
- [15] Štícha, F. (2007). Korpusové statistiky a slovotvorná produktivita. In *Grammar & Corpora/ Gramatika a korpus 2005*. F. Štícha and J. Šimandl (eds.). Praha, Academia, pages 250–257.
- [16] Štícha, F. (2009). Slovotvorná produktivita a gramatičnost: gradační expresivní adjektiva s prefixy pra-, pře- a vele- v současné psané češtině. In *Eslavística Complutense*, 9, pages 145–170.
- [17] Ševčíková, M. (2014). Zjišťování slovotvorné produktivity z korpusových dat: přípony odvozující názvy vlastností. In *Naše řeč*, 97, pages 228–240.
- [18] Baayen, H. R., and Lieber, R. (1991). Productivity and English derivation: a corpus-based study. In *Linguistics*, 29, pages 801–843.
- [19] Baayen, H. R. (1992). Quantitative aspects of morphological productivity. In *Yearbook of Morphology 1991*. G. E. Booij and J. van Marle (eds.), pages 109–149, Dordrecht, Kluwer Academic Publishers.
- [20] Baayen, H. R. (1993). On frequency, transparency, and productivity. In *Yearbook of Morphology 1992*. G. E. Booij and J. van Marle (eds.), pages 181–208, Dordrecht, Kluwer Academic Publishers.
- [21] Baayen, H. R. (1994). Productivity in production. In *Language and Cognitive Processes*, 9, pages 447–469.
- [22] Baayen, H. R., and Renouf, A. (1996). Chronicling The Times: Productive Lexical Innovations in an English Newspaper. In *Language*, 72, pages 69–96.
- [23] Baayen, R. H. (2009). Corpus linguistics in morphology: morphological productivity. In *Corpus Linguistics. An international handbook*. A. Lüdeling and M. Kyto (eds.), pages 900–919, Berlin, Mouton De Gruyter.
- [24] van Marle, J. (1992). The relationship between Morphological Productivity and Frequency: A Comment on Baayen's Performance-Orientated Conception of Morphological Productivity. In *Yearbook of Morphology 1992*. G. E. Booij and J. van Marle (eds.), pages 151–163.