

Online dictionary content and the user needs: The Slovak case

Alexandra Jarošová, Slovak Academy of Sciences

While in classical dictionaries the emphasis was on data (in an academic dictionary, the user often struggles with the problem of information overload), an electronic dictionary focuses on the user's perspective. For the lexicographer this means not merely the need to search for a suitable ergonomic arrangement of dictionary data. Within such approach the very theoretical bases of lexicography are being redefined in the sense that the dictionary is primarily a utilitarian product whose function is to satisfy certain information and communicative needs of the users. At the federated search portal of the L. Štúr Institute of Linguistics, there are available several digitized versions of printed lexicographical works that have been published by this academic institution. The first and the second part of this paper deals with dictionaries in the context of language cultivation and its alternatives. As the prescriptive codification (i.e. dividing linguistic means into standard and non-standard ones) constitutes part of the culture of the users of the Slovak language, the page-views of online dictionaries are very high. However, none of the above referred to Slovak dictionaries available online (neither the descriptive nor the prescriptive ones) sufficiently reflect the most frequent requirements of the users, and the second part of the paper will try to explain why this is the case. The everyday linguistic problems of the users are being dealt with by the "non-dictionary reference genre", namely by the telephone and internet Advisory Services of the Institute of Linguistics. The most frequent types of information required by the users (e.g. questions concerning capitalization, hyphenation, spelling and semantics of neologisms, and declension and conjugation) are discussed in the third part of the paper. Another aim of the third part is to indicate in what way the data collected via the questions addressed to the Advisory Services and the answers provided, together with the three different types of expertise (theoretical morphology, corpus linguistics, and monolingual lexicography), determine the design of the Slovak digital-born Orthographic and Grammatical Dictionary that is at present being created.

Keywords: *user needs, language advisory services, orthographic and grammatical dictionary, prescriptive codification*

1. Introduction

The *Language Advisory Services of L. Štúr Institute of Linguistics* is a popular institution that for four hours daily deals with the language-related questions of the citizens who by phone turn to it for advice. The advisors are also the authors of brief radio contributions addressed to the general public, and they provide answers to the questions that had been sent by e-mail (<https://slovensko.rtv.s.sk/relacie/slovenscina-na-slovicko>). The questions are often aimed at finding out whether the particular word forms part of Standard Slovak ("Is this word correct?"), or which of the two forms of the word is standard/correct. Quite often, the answers are formulated as: "We recommend/do not recommend that you use this word (in rather formal situations), because it is standard/it is not standard, and it is standard/it is not standard, because it

is present/it is not present in the codification dictionary.” However, this strategy cannot be used by linguists when they assess e.g. a new expression borrowed from English. In such case, they try to search for possible suggestions of Slovak equivalents, or for the ways of the grammatical adaptation of the Anglicism. It is evident that such type of answer that does not clarify the rule behind accepting or refusing the word does not seem to be appropriate to all users of the language. They want to understand the “logic” of the phenomenon. On the other hand, it should be noted that the short time available for responses in some cases does not allow for at least a brief explanation.

It stems from the above that a specific trait of the Slovak language-related and linguistic situation – the nurturing of an active relationship with regard to the so-called language cultivation, – based on authoritative prescriptive codification. The concept of cultivating the standard language appeared in the late 1920s in the context of the standardization activities in Czech. It became more famous under the name *Theory of Language Cultivation*, after the *Prague Linguistic Circle* published the seminal texts explaining the attitude of the *Circle* to language intervention into the standard language on the basis of functionalist principles (Havránek & Weingart eds. 1932).

2. Slovak dictionaries in the context of the *Theory of Language Cultivation* and the alternatives of this theory

In Slovakia, the *Theory of Language Cultivation* is developing and modified in connection with the specific features of the particular stage of social development (see Nekvapil 2008 for a broader context).

The traditional understanding of language cultivation is carried out within two aspects. The first aspect of language cultivation deals with the quality of the linguistic usage in public communication and the second aspect focuses on directing the linguistic usage, based on advocating for the valid codification on the part of experts, i.e. the qualified users of the language, and, on the other side, on accepting the valid codification on the part of ordinary users (Ružička 1967; Kačala 1971; Král & Rýzková 1990).

Such understanding also forms the basis of the currently applying and several times amended *Act on the State Language* of 1995 that relies on the existence of the so-called codification manuals, i.e. on academic grammar and on academic dictionaries of three types: on the orthographic-grammatical dictionary as part of *Pravidlá slovenského pravopisu* [Rules of Slovak Spelling] (Považaj ed. 2013, 4th edition), on the orthoepic dictionary as part of *Pravidlá slovenskej výslovnosti* [Slovak Pronunciation Rules] by Král (2009, 2nd edition), and on the one-volume explanatory dictionary *Krátky slovník slovenského jazyka*, henceforth *KSSJ* [Concise Dictionary of the Slovak Language] (Kačala & Pisárčiková & Považaj eds. 2003, 4th edition). The given dictionaries, the material basis of which was being formed in the last part of the 1980s, function as codification manuals within their slightly updated issues. It is the *Ministry of Culture* that is authorised to award the status of codification manual to a particular linguistic publication ([http://www.culture.gov.sk/posobnost-ministerstva/statny-jazyk/kodifikacne-](http://www.culture.gov.sk/posobnost-ministerstva/statny-jazyk/kodifikacne)

[prirucky-c6.html](#)). This is connected with the model of linguistic diversity that differentiates linguistic means into the standard (correct and appropriate) and substandard (inappropriate) ones.

The headword list of the above-mentioned one-volume explanatory dictionary (*KSSJ*) is constituted by a selection in which the preference of the linguistic means created in compliance with the *linguistic system* was applied (Kačala 1994: 102), i.e. the description was made within the framework of the structuralist theoretical platform. In the *Act on the State Language of the Slovak Republic* (1995: paragraph 2, section 3) the Ministry of Culture refers to *regularities/dispositions of the codified form of the State language* [*zákonitosti kodifikovanej formy štátneho jazyka*] and considers these *dispositions* as being the only possible ones: “Any interference into the codified form of the State language in contradiction with its dispositions is inadmissible.”

The delimitation of words into standard and non-standard is system-based (Kráľ 2000: 77–81). Certain formal features of words are considered to be more appropriate from the point of view of the linguistic system of Slovak, hence “more Slovak”. Non-systemic words get disqualified by means of a set of qualifiers. Actually, a certain vicious circle can be identified here. Within the conception of language cultivation, a set of linguistic units presented in the prescriptive codification manual (dictionary) is considered to represent the norm. Such a norm with the features of an ideal has the character of an evaluative measuring device binding for the users of Slovak. In the conception of language cultivation, the implementation of the norm (understood in this way) into the linguistic usage, i.e. the regulatory activities of experts, have approximately the following sequentiality: linguistic system/model – linguistic norm/codification – linguistic culture – language user. Hence, in this traditional chain, the language user as the object of the impact occurs on the last place. When creating the *KSSJ*, in the centre of attention was not the user but the linguistic system. The application of the model of standardness/non-standardness causes a certain kind of communication problems and, at the same time, generates manuals that seemingly resolve these problems. Even in the case of a frequently occurring word that is used in formal situations, the users are not sure about its standard character.

In Slovak linguistics, however, there also exists a socio-linguistic and linguistic-pragmatic alternative to this conception based on the idea that language is to be interpreted from the position of its user in discourse, i.e. from the position of an ordinary user’s linguistic consciousness (Dolník 1996; Dolník 2010). In this conception, the norm forms part of the complex of standardizations, it represents the usage by the majority of speakers, and it is marked by natural variability. According to the protagonists of this approach, only this natural norm should be codified in linguistic manuals. Instead of language cultivation in the sense of authoritative regulation of practical linguistic usage, there should be applied linguistic management, i.e. linguistic advisory services based on the current needs of the language users.

The representatives of the traditional understanding of the concept of language cultivation in the 1990s declared that the state of linguistic culture in the sense of the quality of linguistic usage is catastrophic.

The representatives of the alternative sociolinguistic position expressed the conviction that linguistic usage is not in a catastrophic state, as many variants of the norm can be considered as being standard variants of the norm and not as linguistic mistakes. Within this linguistic approach, the declared transfer of interest to the language user has been carried out above all in

the form of thinking about an appropriate theoretical model of language which would replace the structuralist model of the “linguistic system” that ignores the mental reality of the speakers. The attitudes of the speakers were rhapsodically investigated with the help of questionnaires in which sociolinguists formulated questions concerning phenomena which they themselves considered to be problematic. The problems were being raised by linguists, not by users.

The *Rules of Slovak Spelling (PSP)* and the *KSSJ* (each containing 61 thousand entries) are the products of lexicography aimed at satisfying the communication needs of the wide public, i.e. the products of the so-called non-scholarly (though still academic) lexicography.

The new corpus-based *Slovník súčasného slovenského jazyka*, henceforth *SSSJ* [Dictionary of the Contemporary Slovak Language], is a representative of scholarly lexicography fulfilling both the cognitive as well as the communicative function. Up to now, the third volume of *SSSJ* (out of the eight planned ones) has been published (Jarošová ed. 2015). The *SSSJ* that is being published is aimed at describing the real linguistic norm, i.e. the regularly used, conventionalized language items. Thus what is described is not an idealized norm as a construction created by linguists, but the norm as a phenomenon based on language usage. This particular concept of the norm (“a complex of grammatical and lexical means (structural as well as non-structural), which are regularly used”) was developed within the functionalist theoretical framework of *Prague school* (Havránek 1932: 33). The real norm contains a large number of lexical and morphological variants, as well as foreign words the degree of adaptation of which varies. The above resulted in the fact that the function of the dictionary is to be a detailed description of the lexis based on the generalization of a large number of empirical data (we have at our disposal an in-house corpus *Omnia Slovaca* containing 4.9 milliard tokens) in compliance with the theoretical model of the language. For preventing the risk of an overload, lexicographers are assisted by the very useful theoretical model by Hanks (2013) built on the opposition of “norms (conventional uses of expressions) and dynamic/ad hoc exploitations of norms”. The object of lexicographical description should be constituted by the conventional usage of linguistic means.

It is also necessary to take into consideration the codificational continuity, and to a certain extent respect the results of the previous lexicographical agenda in the sphere of delimiting the means into standard and non-standard. The lexicographers involved in the preparation of the dictionary aim at improving the existing reduced model of the lexical meaning based on the functional-structuralist basis (Jarošová 2018a; Jarošová 2018b). We are expanding this model by including the concepts presented within other theoretical and methodological frameworks, such as sociolinguistics, linguistic pragmatics, cognitive linguistics and corpus linguistics (all of them departing in some respects from structuralism and, in other aspects, being complementary to it). The extended model of lexical meaning constitutes a certain synthesis of the given theoretical frameworks and, at the same time, represents a reflection of three language constituents:

1. The social constituent is present in the form of the consideration of the communicative functions of utterances, of the naming functions of lexical units, of functional styles and registers, of language norms, and of situational contexts.

2. The psychic component presents the consideration with regard to the prototype effect, the abolition of boundaries between linguistic meaning and other parts of cognitive content.

3. Thanks to the structural-systematic component, a description of the paradigmatic and syntagmatic behaviour of words can be carried out, and an inventory of formal-content units and categories (lexemes, lexias, word-formative and grammatical structures) can be provided.

Our ambition as lexicographers is nothing less than a word-centered description of the whole language, but we have not asked the question of how the user can extract the necessary information from this construct. Neither did lexicographers in preparing this dictionary primarily base their work on the real needs of the user, but they dealt with their own scholarly linguistic agenda.

3. A need of a new user-oriented manual in the context of new technologies

Electronic media enable us to see the content-related as well as the presentation-related component of the dictionary in a new light. Corpus-based methods as tools with the help of which we look at raw textual data, have enabled us to observe the recurrent patterns of language, the scale-based and the continual character of linguistic meaning, but also of linguistic form. The corpus has relativized the langue-parole opposition. This langue-parole continuum has to be interpreted in some way, as well as discretized and selected, with the aim of creating units and categories. Linguistic data are not self-evident. They are a result of hard analytical work. Connected with the above is also the aforementioned building of a new model of lexical meaning and of its application variant, i.e. of the dictionary entry.

As soon as we have a sufficient amount of data, we face the problem of designing the dictionary data, which is pointed out by papers targeted at user-research projects (Müller-Spitzer et al. 2012; Tiberius & Müller-Spitzer 2015). While in classical dictionaries emphasis was placed on data (in academic dictionaries, the user often struggles with the problem of information overload), electronic dictionaries focus on the user's perspective. Authors of dictionaries were at first fascinated by the possibilities offered by a dictionary as an original electronic product. Where are the limits of academic dictionaries? From the point of view of the scholarly completeness and the volume of "storage databases", no limits can be established. Nevertheless, Lew (to be published) points out that from the aspect of appropriateness for the user and of the possibilities of handling the large volume of data, the restrictions are strong. Dictionaries that from the very beginning are created as electronic products, have extensive possibilities of dynamically representing their contents. At the disposal are numerous search fields enabling interactive selection, multimedia attachments to entries, hypertext links to other linguistic sources, e.g. to corpora of texts, to other relevant dictionaries, etc. Dictionaries can have the form of a database enabling searching via the particular parameters (e.g. via the data on the origin of the item), but can also enable full-text searching. We believe that namely the possibilities offered to contemporary information technologies have stimulated the situation within which the user is in the centre of attention. This is related not only to searching for an appropriate ergonomic organization of lexicographical data (Müller-Spitzer et al. 2012). The very theoretical basis of lexicography is being redefined in the sense that a dictionary is above all a utilitarian product the function of which is to satisfy certain information needs of its users. According to this conception entitled *Function Theory of Lexicography* (Bergenholtz & Tarp 2003), the efforts of

lexicographers have to concentrate above all on determining these needs on the basis of identifying the specific groups of users and specific situations within which the users turn to the dictionary. The representatives of functional lexicography, referred to also as *Aarhus School of Lexicography*, stress the fact that in the situation of using an electronic dictionary, the user can and should be an active user: “the essential problem of dynamic data does not reside in the storing of the data in the database, but in finding ways for presenting the data dynamically to the users [...] to fit in with the needs of the user in a given user situation (Andersen & Nielsen 2009: 360)”. Thanks to information technologies, the needs of the users are satisfied with the help of an interactive offer. The representatives of *Aarhus School*, pointing out namely the said utilitarian character and the strong technological component (Tarp 2012) consider lexicography to be a discipline including all reference manuals, and to be part of information science. Although lexicography borders on information technology (here we would like to point out the fluid borderline between general dictionaries, specialized dictionaries, encyclopaedic dictionaries and encyclopaedias themselves), it is still deeply rooted in linguistics.

However, in contrast to *Aarhus School*, we understand lexicography in a narrow sense of the word, hence as a technology of the presentation of language-related data for the purposes of satisfying the information needs connected with the communication of the user, hence not as a discipline including all the referential handbooks. Technology has to be indispensably based on scholarly findings, those being provided above all by linguistics, and at present, to a large extent, also by information science. In my opinion, it would be a better solution to use the term *lexicography* to refer to language-oriented dictionary manuals, while the wider area of reference works could be called a different way, e.g. *informgraphy*. Hence, we do not refuse the interdisciplinarity and the continual character of information/reference tools.

Slovak lexicography has so far been using the possibilities of electronic media and linguistic technologies only to a lesser extent. This applies above all to the area of the presentation component of the dictionary. At the federated search portal of the L. Štúr Institute of Linguistics (<http://slovniky.juls.savba.sk/>), there are available several digitized versions of printed lexicographical works that have been published by this academic institution. The string searched for can be constituted by the whole lemma or its parts.

As prescriptive codification forms part of the culture of the users of the Slovak language, in the case of online dictionaries the page-views are very high. The Slovak Republic has 5.44 million inhabitants, out of which 3.75 are at productive age. In the year 2017, the dictionary portal had 460 thousand users and 2.2 million page-views.

However, numbers do not testify to the quality of the consultation sources. *KSSJ* and *Spelling dictionary* have a too reduced content and they are not regularly updated. Neither do they fulfil their principal function of refining the linguistic culture of the user, as this function stems from the problematic premise that idealized norm can be transferred into practical usage with the help of a dictionary.

The scholarly dictionary (*SSSJ*) contains many highly structured data and shortenings denoting linguistic categories. It is difficult to get oriented within this diversified space. It is not easy to find a particular item of information which would fulfil the “punctual information needs” (using the terminology of Tarp 2012: 101–112), although it is probable that the information is contained in the dictionary.

How is this offer viewed by the user? The statements made on the webpage of amateur linguistic advisors that is called *Milujeme slovenčinu* [We Love Slovak] (<http://www.milujemeslovincinu.sk/>) testify to the fact that there have appeared active users who often hold critical opinions with regard to our dictionaries: “The Institute of Linguistics should exert activities directed at the nation – they should get out of the shells of their offices where, with their noses dipped into academic papers and dictionaries, they do not see that, slow but sure, they are becoming an isolated isle of their own. People have many questions which have probably been answered somewhere, but, unfortunately, hardly anyone has enough abilities and patience to search for answers in complicated expert manuals. Some more integrated activities of the Institute of Linguistics would be very welcome. So far it is only me alone who is attempting to do that via this community web page *Milujeme slovenčinu*” (Zbínová 2012).

From our point of view it is important to state the reason which has led to the founding of the unofficial advisory portal: the manuals that offer the official dictionary portal of the Institute of Linguistics are too complicated and too “specialized” for ordinary and quick searching.

On the other hand, not only criticism is addressed to the dictionary, and its authors receive a considerable number of positive responses. Users appreciate the extended manner of definitions and the adequate presentation of exemplifying collocations and sentences.

The discussion forums focusing on language usage provide to linguists the possibility of gaining a number of immediate items of information on the attitudes and the needs of the users. Such information has to be inspected closely, as a reasonable innovation of the existing dictionaries and the designing of online dictionaries of the new type cannot do without such user-generated content (Lew 2014; Tarp 2015).

The everyday linguistic problems (“punctual information needs”) of the users are being managed by the non-dictionary “reference genre”, i.e. by the above-mentioned telephone-operated Advisory Services of L. Štúr Linguistic Institute. Until the year 2013 there existed advisory services offered by telephone or by mail. The linguists working at the advisory services handled about 10 thousand questions yearly. Since 2013 the project of internet advisory services was started, based on the principle of the database containing 5532 frequently asked questions together with answers to them. In the year 2017, the internet advisory services numbered 130 thousand real users and 300 thousand page-views. Within the telephone advisory services about 6000 answers were given. Functioning in a limited regime are also advisory services by means of letters that tackle about 600 letters yearly.

This situation is not favourable for users in spite of the fact that the advisory activities are of such a multi-genre character. Codification manuals are under-dimensioned from the information point of view. Moreover, in the case of the *SSSJ*, the user struggles with the problem of information overload, and, above all, with the aged presentation technology which does not make use of the possibilities of electronic media. From our point of view, a model for electronic adjustment of a paper version dictionary to be followed is *Den Danske Ordbog* [The Danish Dictionary] (https://ordnet.dk/ddo_en; for more information see Trap-Jensen 2010).

The internet advisory service can satisfy only a small number of those interested in linguistic consultations, and the telephone of the advisory services is often busy, hence difficult to reach. The time has come for creating a new alternative dictionary inspired by the real needs of real users.

How can we find out what the user is searching for? In trying to find the answer, we can be inspired by Slovenian linguists who decided to use the information from the linguistic advisory services on the internet and have created a very well designed bottom-up categorization (Arhar-Holdt et al. 2017: 3). At the first four places we can find the following questions: “Is this word correct or not?” (this question can actually also mean the verification of the fact whether linguists consider the particular word to be part of the Slovenian lexis at all); “Which of these options is better?”; “How is this word declined?”, “What does this word mean?” This is, in fact, the type of information searched for also by the users of the popular dictionary app *Svenska Akademiens Ordlist* [the Swedish Academy Glossary]: “about 57% of respondents mostly use the app to check spelling or meaning [...], [a]bout 54% use it to check ‘if the word is included in the glossary’ [...], 53% look for inflection” (Holmer et al. 2015: 364). It is evident that in addition to the question concerning the meaning, the consultation needs of the users are satisfied by the type of dictionary which can be called orthographic and grammatical.

Let us have a look at the information provided by the Slovak advisory sources. The typology of the issues made available to us by those working in *Language Advisory Services* represents a good source of empirical and expert information (Hrubaničová 2017).

1. From the aspect of the frequency of the user questions, the first places are taken by those ones related to punctuation (often a comma before a conjunction), and capitalization (candidates for gaining the status of proper names from the area of institutions, manmade structures, governmental matters, historical events and special occasions; common name candidates from the area of brand names), then hyphenation and shortening (from the area of compounds and other multi-word units). Evidently, the official *Rules of Slovak Spelling* dealing with the rules of using punctuation marks and of naming units having a specific function (e.g. for proper nouns or compounds) are not formulated clearly and unambiguously. Finally, there are the questions concerning the spelling and the normative status of neologism (e.g. Slovak derivatives so far not included in the dictionary, new foreign words, these often being terms).

2. Another frequent problem is represented by the pronunciation of Slovak words with regard to palatal consonants (palatalized consonants which have also their non-palatalized correlates are characteristic of the standard variety, but they are not present in Eastern Slovak and West Slovak dialects).

3. Next comes the meaning of neologisms.

4. Formal morphology

(a) Declension of nouns

(b) Variants of case endings

(c) Morphological case of nouns and shortenings

(d) Declension of foreign proper nouns.

It has to be born in mind that prescriptive codification (delimitation of linguistic means into standard and non-standard) constitutes part of the culture of the users of Slovak. That is why we hold the opinion that the users will also welcome the presentation of usage labels (colloquial, literary, poetic, journalese, administrative, official, specialized/technical, professional; regional, slang, substandard; expressive, pejorative, ironic, familial, facetious, rude, vulgar; rare; archaic, obsolete). It will be necessary to analyze again the concept of the standard and the notional

content of special normative labels with prohibitive function – incorrect, inappropriate (cf. also Šipka 2016).

The general conclusions that have been drawn here on the basis of the data collected by the *Language Advisory Services* thanks to the expertise of this institution can be formulated in the following way. The basis of the new *Ortograficko-gramatický slovník*, henceforth *OGS* [Orthographic and Grammatical Dictionary] will be constituted by the list of validated entries from the existing explanatory dictionaries extended by neologisms and the selected types of proper names which cause problems from the point of view of the usage of capital initial letters in them (multiword proper names), or from the point of view of their declension (foreign surnames).

With the help of the methods of corpus linguistics, we have extracted the above-mentioned neologisms from the balanced sub-corpus (313 441 150 tokens) of the *Slovak National Corpus prim-6.1-public*. The list consists of 17,000 lemmas that so far have not been lexicographically processed. One of the stages of creating the list of neologisms was manual lemmatization of the forms not recognized by the morphological analyzer. This operation also provided us with information on the unregistered morphological variants of entries that had already been part of the “old” list. The innovative works based on analyzing Slovak morphology have produced a tool for distinguishing whether the unregistered form means a mistake or a “systemic variant” (Sokolová 2007; Sokolová 2012). A variant is the result of the simultaneous functioning of two or more factors (e.g. the masculine suffix *-tel'* typically denoting a person is in collision with the meaning of the word *delitel'* (divisor) denoting an object; *aids* as to its spelling ranks into the non-palatalized declension pattern, and as to pronunciation into palatalized declension pattern). This phenomenon causes the coexistence of variant endings, e.g. *delitel'u/delitel'ovi* in DatSg, *delitele/delitelia* in NomPl; *aidse/aidsi* in LocSg. The character of the variantness of verbs is interesting, too. Some conjugation types are “strong” (they do not have alternations in the root, and they have unequivocally predictable endings within the extent of their whole paradigm). For example, conjugated according to the three strongest conjugation patterns (*chytat'* (to catch), *pracovat'* (to work) and *robiť* (to do) are 80 % of Slovak verbs. Their prototypical character causes that verbs from the “weaker” types start to be conjugated according to them. This is an area that generates forms with variant endings, e.g. *klzat'* ‘slide’: *klže* (Prs3Sg), *klžu* (Prs3Pl), *klž!/klžaj!* (Imp), *klžal* (3SgPtt), *klžuc/klžajúc* (Transgressive), *klžuci/klžajúci* (ActPrsPt). Within the process of analysis we deal with all the variant forms, but we accept and present only those that arise as a result of regular cross-conjugational interference. Some variant forms arise on the basis of the penetration of dialectal endings, e.g. the form *bere* (takes) that is widespread in the West-Slovak and the East-Slovak regions has its standard counterpart in *berie*. However, within the grammatical data concerning the verbs we do not present forms with dialectal endings. Some variants are distinctly more frequent and they constitute the norm. The variants with a very small occurrence constitute marginal realizations of the systemic potential. We have to resolve the question whether we should include such variants into the description, or whether their inclusion should be delimited by some percentage. Within morphological variantness we set the limit for presenting the minority variant at ≥ 10 % from the overall occurrence of the particular form. This is based on understanding the norm as usage by the majority, and 90 %, beyond any doubt, represents a majority. All “interferencing” variants the occurrence of which is below 10 %, are systemic, hence correct, but they do not constitute

part of the majority norm. In case of each variant, in addition to frequency we pay attention to the variedness of the sources of occurrence, as well as their value (an older text, e.g. from the period between the 1950s and 1980s, versus a contemporary text, i.e. from the period from the 1990s up to now; an original or a translated text; internet discussion, or a blog versus the official page of the institution). These factors increase or decrease the weight of occurrence.

In case of orthographic variantness a rather problematic group is constituted by Anglicisms. These words undergo a process of adaptation that generates numerous variants. For example, along with the frequently occurring original form *hacker* there also occurs a certain number of variants having a considerable lower frequency of occurrence: *haker* (hybrid form), *heker* (fully adapted form based on pronunciation), and *hecker* (a hybrid form). For inclusion into the *SSSJ*, in addition to the original form *hacker* we selected the variant *heker*. Although it does not rank as the second most frequent one, its selection is supported by a number of factors: its graphical form is the simplest, it is not a hybrid, and it has manifested extensive derivational potential (there have been formed verbs as *hekovat'* and *heknút'*, the adjective *hekerský*, the abstract noun *hekerstvo*, the adjectivized particles *hekovany'* and *heknuty'*, and the adjective *hekovatel'ný*). In addition, the occurrence of the form *heker* is increasing and it is used by prestigious periodicals. Hence, the basis of the decision-making is formed by frequency, but other factors are also considered. Not all Anglicisms undergo a process of adaptation within which the graphical form of the verb gets changed. Marked by a high degree of resistance are Anglicisms from the sphere of music, and, on the contrary, Anglicisms from the sphere of sport are often adapted.

At the first stage, the dictionary will present the graphematic form of the lemma and the lemma's grammatical and pronunciation data in the format that has been set for entries in the *SSSJ*. Hence, the entries will look like partial entries in the *SSSJ*. In this version the dictionary will present the identifying morphological forms including the nouns and verbs into the particular declension and conjugation class. Part of these forms constitute neuralgic points of the paradigm as they occur in variants and the users feel uncertain about them. In the previous dictionaries, to the detriment of the situation, many of these neuralgic points were not presented.

Homonyms will be accompanied by a brief explanation, and an explanation will also be produced for neologisms that so far have not been lexicographically processed.

Hence, the entry will contain the following data:

- (1) Lemma
- (2) Homonym number
- (3) Pronunciation (where needed)
- (4) Variant spelling (where needed)
- (5) Variant inflection (where needed)
- (6) Grammar and spelling remarks (where needed)
- (7) Declension forms displayed in a table
- (8) Usage labels ("qualifiers")
- (9) Meaning (in the case of homonyms and neologisms).

The typological range of the related usage labels is as follows:

- (1) Subject domain (102 symbols for the particular scientific and technological fields)

- (2) Character of the word from the point of formality/informality of communication (e.g. a colloquial word, an official denotation, etc.)
- (3) Appurtenance of the word to a particular sociolect (e.g. youth slang)
- (4) Appurtenance to levels of style (e.g. literary word, poeticisms, journalistic expression, biblical word)
- (5) Frequency of the word (rare word)
- (6) Attitude – Attitude to the communicated content (e.g. a derogatory word, an ironical word)
- (7) Aspect of the prescribed normativeness (a not recommended word, substandard word, incorrect word)

Only three of the above labels exclude the word from the standard language: substandard, slang and incorrect. The labelling of a particular linguistic means as incorrect had been worked out within the codification agenda of the previous dictionaries and, as a rule, it concerned the Bohemisms which had only undergone a phonological adaptation process (diphthongization, shortening of vowels and palatalization of the consonants *d*, *t*, *n*, *l* at the particular places), but their roots, affixes and manners of consonantal alternation have preserved their Czech character. Based on well-grounded reasons, we did not accept part of this agenda. On the other hand, with the aim of – also well-grounded – codification continuity, we took over part of the agenda also into the *SSSJ* and *OGS*.

In *KSSJ*, some words that were labelled as incorrect, or their Slovak equivalent was labelled by the qualifier *správne* ‘correct’, abbrev. as *správ.*, are labelled in *SSSJ* by a different qualifier, e.g. by the qualifier *hovorový výraz* ‘colloquial expression’, abbrev. as *hovor.* in the entries *hmoždinka* ‘coak’, *čípok* ‘hip (as medication)’; with a reference to *vhodnejší výraz* ‘a more appropriate expression’ in the case of the entries *jedálniček* ‘menu’, *behom* ‘(by) running’, *klud* ‘relaxation’, ‘coolness’, ‘standstill’, ‘peace’, *čiastka* ‘(financial) sum’; with the label *regionalizmus* ‘regionalism’, abbrev. as *region.* in the entry *krecht* ‘potato bank (for storage)’; the entry *nezávadný* ‘without any flaws’, ‘unobjectionable’ is labelled as *odborný výraz* ‘technical expression’, and the entry *obora* ‘game preserve’ has no label, i.e. it is considered as being neutral along with its synonym *zvernica*. For these words we do not have any semantically and pragmatically completely equal and sufficiently used Slovak equivalent, and the equivalents suggested, in spite of the several decades since then, have not entered into usage (they differed in some semantic aspects or in collocability). These Bohemisms enrich the sets of synonyms and we consider them to be part of the Slovak lexis. Both in *SSSJ* and *OGS* we label as incorrect a small number (107 out of 155,000 entries) of frequent unadapted Bohemisms that are parallelly used with the semantically completely identical and current Slovak equivalents, e.g. in the entry *bojácný* reference is made to its correct Slovak counterpart *bojazlivý* ‘fearful’, and similarly processed are the pairs of words *čidlo* → *snímač* ‘sensor’, *dielčí* → *čiastkový* ‘partial’, *dosažiteľný* → *dosiahnuteľný* ‘achievable’, *jaderný* → *jadrový* ‘nuclear’, *lehátko* → *ležadlo* ‘deckchair’, *krunier* → *pancier* ‘armouring’, *lomítko* → *lomka* ‘slash’, *nahorklý* → *horkastý* ‘bitterish’, *ožehavý* → *pálčivý* ‘poignant’, etc. Some words concerned are internationalisms the orthography of which has become stabilized in a certain manner, and by educated people the modification of this manner is not considered to be a variant, but a mistake. For example, the form *gramofón* (record player) is considered as being standard and the form *gramafón* as a

mistake. Similarly perceived is *antedatovať* ‘antedate’ as against the non-standard *antidatovať*, *kontroverzný* ‘controversial’ as against *kontraverzný*, *ekvipáž* ‘(horse-drawn) equipage’ as against *ekipáž*, *percento* ‘percentage’ as against *procento*, etc. The areas where “incorrect” Bohemisms get used are neutral, formal or specialized communication. We explain to the users the reference to the “correct” Slovak equivalent as a recommendation not to use the particular Bohemism in public communication.

In contrast to such lexical Bohemisms that by linguists are considered to be unrecommendable counterparts to standard lexical units, the substandard involves a set of borrowings that are widely used in general informal communication. These are mostly Germanisms (some borrowed through Czech) and Bohemisms. There also occur Anglicisms, but those are mostly used in youth slang and in professional slang, i.e. in informal group communication. Orthographically, the words borrowed from German and English are based on Slovak pronunciation, and not on their original graphical form. However, they usually comply with the Slovak grammatical paradigms (*cušpajz* ‘sauce’, *dunst* ‘idea’, ‘steam’, *fofr* ‘haste’, *hexenšus* ‘put-out back’, *kasírovať* ‘collect’; *líbling* ‘sweetheart’; *párty* ‘party’, *lúzer* ‘loser’, *ofis* ‘office’. Czech words have been borrowed in their original graphical form with characteristic Czech roots (*čumák* ‘muffle’, *ješitný* ‘vain’, *kecy* ‘yak’, *bulíkat* ‘cheat’) and affixes (*kutil* ‘bricoleur’, *mľaďas* ‘youngster’, *nastojáka* ‘while standing’). These lexical units tend to be used with the aim of their differentiation from the standard norm, and they often have an expressive-evaluative feature.

4. Conclusion and future work

While in classical dictionaries emphasis was placed on data, the electronic dictionary focuses on the user’s perspective. This means not only searching for a suitable ergonomic arrangement of dictionary data. Actually, the very theoretical basics of lexicography are being redefined in the sense that the dictionary is primarily a utilitarian product whose function is to satisfy certain information and communicative needs of the users.

The *Orthographic and grammatical dictionary* (<http://lex.juls.savba.sk/>) is designed as a dictionary which, to a larger extent than it was until recently, takes into consideration the needs and the interests of the ordinary user, i.e. not only of the needs of a professional who uses language as a tool for his or her work. In making the *OGS*, we use the method of joining the know how from several areas: from language advisory services, from theoretical morphology, from corpus linguistics, and from monolingual lexicography. Tarp (2014) gave a very instructive description of this method and denoted it as a *functional method*. We completed the headword list with basic grammatical information in 2016 (this first version is on the web). Till the end of May 2019, we dealt with verifying and filling in the relevant orthographic and morphological data of the headwords in the dictionary. A lot of attention has been paid above all to verifying the variant forms in the corpus, with the aim of determining the boundaries of inflectional variability and differentiating variants from marginal deviation. The dictionary is an autonomous product and, at the same time, it forms the basis for the headwords of the following volumes of *SSSJ*. *OSG* contains 155,000 headwords and some of them are not listed in the so-far published

volumes of *SSSJ*, as they are either new, or they are too narrowly specialized. It also contains some additional morphological data (e.g. variants) motivated by the increased data in the corpora, hence it provides more reliable information about the grammatical behaviour of the words. From the user's point of view, continuous and frequent updating in such dictionaries constitutes a huge advantage. In future, we intend to add into the entry data about the whole paradigm in the case of nouns, adjectives, pronouns, and numerals, as well as data about the linguistic register, and incorporate brief explanations into the headwords representing specialized terms and neologisms. We hope that on the basis of its topicality and reliability the dictionary will become popular and authoritative among its users also without its having the status of a codification manual.

Abbreviations

KSSJ *Krátky slovník slovenského jazyka* [Concise Dictionary of the Slovak Language]
OGS *Ortograficko-gramatický slovník* [Orthographical and Grammatical Dictionary]
SSSJ *Slovník súčasného slovenského jazyka* [Dictionary of the Contemporary Slovak Language]

References

- Andersen, Birger Enevold & Nielsen, Sandro. 2009. Ten key issues in lexicography for the future. In Bergenholtz, Henning Nielsen & Sandro Tarp, Sven (eds.), *Lexicography at a crossroads: Dictionaries and encyclopedias today, lexicographical tools tomorrow*, 355–363. Bern: Peter Lang.
- Arhar Holdt, Špela & Čibej, Jaka & Zwitter Vitez, Ana. 2017. Value of language-related questions and comments in digital media for lexicographical user research. *International Journal of Lexicography* 30(3). 285–308.
- Bergenholtz, Henning & Tarp, Sven. 2003. Two opposing theories. On H.E. Wiegand's recent discovery of lexicographic functions. *HERMES - Journal of Language and Communication in Business*, 16(31). 171–196. (<https://tidsskrift.dk/her/article/view/25743>) (Accessed 2019-02-18)
- Dolník, Juraj. 1996. K analýze teórie jazykovej kultúry. [Toward an analysis of language cultivation theory.] *Slovenská reč* 61(5). 282–291.
- Dolník, Juraj. 2010. *Teória spisovného jazyka so zreteľom na spisovnú slovenčinu*. [A theory of standard language with regard to standard Slovak.] Bratislava: Veda. (http://www.juls.savba.sk/attachments/pub_teoria_spisovneho_jazyka/spijaz.pdf) (Accessed 2019-02-18)

- Hanks, Patrick. 2013. *Lexical Analysis. Norms and Exploitations*. The MIT Press.
- Havránek, Bohuslav. 1932. Úkoly spisovného jazyka a jeho kultura. [The tasks of the standard language and its culture.] In Havránek, Bohuslav & Weingart, Miloš (eds.), *Spisovná čeština a jazyková kultura*. [Standard Czech and Language Cultivation.], 32–84. Prague: Melantrich.
- Havránek, Bohuslav & Weingart, Miloš (eds.). 1932. *Spisovná čeština a jazyková kultura*. [Standard Czech and Language Cultivation.] Prague: Melantrich.
- Holmer, Luise & Hult, Ann-Kristin & Sköldbberg, Emma. 2015. Spell checking on the fly? On the use of a Swedish dictionary app. In Kosem, Izvor & Jakubiček, Miloš & Kallas, Jelena & Krek, Simon (eds.), *Electronic lexicography in the 21st century: Linking lexical data in the digital age*. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom, 356–371. Ljubljana: Trojina, Institute for Applied Slovene Studies. (https://elex.link/elex2015/proceedings/eLex_2015_23_Holmer+Hult+Skoldberg.pdf) (Accessed 2019-02-18)
- Hrubaničová, Ingrid. 2017. Typy otázok a odpovedí v listovej jazykovej poradni v r. 2005 – 2015 a z nich vyplývajúce možné inovácie v poradenskej činnosti. [Types of questions and responses occurring in letters sent to and from the Language Advisory Services (from the years 2005 to 2015) and the possible innovations in the language advisory activities.] (Paper presented at the conference *Jazykové problémy, ktorých riešenie sa ťažko hľadá*. [Linguistic problems whose solution is difficult to find.] Bratislava, 28–29 November 2017.)
- Jarošová, Alexandra. 2018a. Špirála poznania alebo funkcionalizmus a kognitivizmus (v lexikológii). [The knowledge spiral or functionalism and cognitivism (in lexicology) through the lens of two linguistic jubilees.] *Slovenská reč*. 83(2). 123–142. (<http://www.juls.savba.sk/ediela/sr/2018/2/sr18-2.pdf>) (Accessed 2019-02-18)
- Jarošová, Alexandra. 2018b. Pohľad na pomenovanie cez prizmu teoretických rámcov a slovníkového hesla. [Looking at a naming unit through the prism of theoretical frameworks and dictionary entry.] *Jazykovedný časopis* 69(3). 277–301.
- Kačala, Ján. 1971. Teória a prax v jazykovej kultúre. [Theory and practice in language cultivation.] *Kultúra slova* 5(3). 65–67.
- Kačala, Ján. 1994. *Slovenčina – vec politická?* [Is the Slovak language a political matter?] Martin: Matica slovenská.

- Kráľ, Ábel. 2000. Odkiaľ vzali Slováci spisovnú slovenčinu? [Where did the Slovaks take their standard language from?] *Slovenská reč* 65(2). 71–85.
- Kráľ, Ábel & Rýzková, Anna. 1990. *Základy jazykovej kultúry*. [Basics of language cultivation.] Bratislava: Slovenské pedagogické nakladateľstvo.
- Lew, Robert (in press). Space restrictions in paper and electronic dictionaries and their implications for the design of production dictionaries. In Bański, Piotr & Wójtowicz, Beata (eds.), *Issues in Modern Lexicography*. München: Lincom Europa. (http://wa.amu.edu.pl/~rlew/pub/Lew_space_restrictions_in_paper_and_electronic_dictionaries.pdf) (Accessed 2019-02-18)
- Lew, Robert. 2014. User-generated content (UGC) in online English dictionaries. In Abel, Andrea & Klosa, Annette (eds.), *Der Nutzerbeitrag im Wörterbuchprozess: 3. Arbeitsbericht des wissenschaftlichen Netzwerks 'Internetlexikografie'*. (OPAL – Online publizierte Arbeiten zur Linguistik 4), 8–26. Mannheim: Institut für Deutsche Sprache. (<https://pub.ids-mannheim.de/laufend/opal/pdf/opal2014-4.pdf>) (Accessed 2019-02-18)
- Müller-Spitzer, Carolin & Koplenig, Alexander & Töpel, Antje. 2012. Online dictionary use: Key findings from an empirical research project. In Granger, Sylviane & Paquot, Magali (eds.), *Electronic lexicography*, 425–457. Oxford: Oxford University Press.
- Nekvapil, Jiří. 2008. Language Cultivation in Developed Contexts. In Spolsky, Bernard & Hult, Francis M. (eds.), *The handbook of educational linguistics*, 251–265. Oxford: Blackwell Publishing Ltd.
- Sokolová, Miloslava. 2007. *Nový deklinačný systém slovenských substantív* [New declension system of Slovak nouns]. Bratislava: Veda. (http://www.juls.savba.sk/ediela/novy_deklinacny_system/) (Accessed 2019-02-18)
- Sokolová, Miloslava. 2012. Vymedzenie konjugačných tried, vzorov, podvzorov a ich variantov v slovenčine. [Delimitation of conjugation classes, patterns, sub-patterns and their variants in Slovak.] In Buzássyová, Klára & Chocholová, Bronislava & Janočková, Nicol (eds.), *Slovo v slovníku: aspekty lexikálnej sémantiky – gramatika – štylistika (pragmatika)*. [Word in dictionary: Aspects of lexical semantics – grammar – stylistics (pragmatics).] 149–158. Bratislava: Veda. (https://www.juls.savba.sk/attachments/pub_slovo_v_slovníku/slovo_v_slovníku.pdf) (Accessed 2019-02-18)
- Šipka, Danko. 2016. Exclusion labels in Slavic monolingual dictionaries: Lexicographic construal of non-standardness. In *Colloquium: New Philologies* 1(1). 1–17.

- Tarp, Sven. 2012. Theoretical challenges in the transition from lexicographical p-works to e-tools. In Granger, Sylviane & Paquot, Magali (eds.), *Electronic lexicography*, 107–118. Oxford: Oxford University Press.
- Tarp, Sven. 2015. Detecting user needs for new online dictionary projects: Business as usual, user research or ...? In Tiberius, Carole & Müller-Spitzer, Carolin (eds.), *Research into dictionary use/ Wörterbuchbenutzungsforschung: 5. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“*. (OPAL - Online publizierte Arbeiten zur Linguistik 2). 17–27. Mannheim: Institut für Deutsche Sprache. (<https://pub.ids-mannheim.de//laufend/opal/pdf/opal15-2.pdf>) (Accessed 2019-02.18)
- Tiberius, Carole & Müller-Spitzer, Carolin. 2015. Introduction. In Tiberius, Carole & Müller-Spitzer, Carolin (eds.), *Research into dictionary use/ Wörterbuchbenutzungsforschung: 5. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“*. (OPAL - Online publizierte Arbeiten zur Linguistik 2), 3–5. Mannheim: Institut für Deutsche Sprache. (<https://pub.ids-mannheim.de//laufend/opal/pdf/opal15-2.pdf>) (Accessed 2019-02-18)
- Trap-Jensen, Lars. 2010. One, two, many: Customization and user profiles in Internet dictionaries. In Dykstra, Anne & Schoonheim, Tanneke (eds.), *Proceedings of the XIV EURALEX International Congress*, 1133–1143. Leeuwarden/Ljouwert: Fryske Akademy – Afûk.
- Zbínová, M. (2012). Prečo vlastne máme chcieť písať bez chýb? [Why should we actually aim at writing without mistakes?] (online) (<http://www.milujemeslovincinu.sk/nazory/preco-vlastne-mame-chciet-pisat-bez-chyb/>) (Accessed 2019-02-18)

Internet sources (Accessed on 2019-02-18)

- Act of Parliament Dated November 15, 1995 on the State Language of the Slovak Republic*: <http://www.culture.gov.sk/vdoc/462/an-act-of-parliament-on-the-state-language-of-the-slovak-republic--1ab.html>
- Slovenčina na slovíčko!* [Drop in for a chat with Slovak!] (the archive of language-related radio contributions): <https://slovensko.rtv.slovensko.sk/relacie/slovencina-na-slovicko>
- Den Danske Ordbog* (the online Danish dictionary): https://ordnet.dk/ddo_en
- Slovenské slovníky* [Slovak dictionaries] (the dictionary portal of the L. Štúr Institute of Linguistics): <http://slovniky.juls.savba.sk/>
- Jazyková poradňa Jazykovedného ústavu Ľ. Štúra* [Advisory Services of the L. Štúr Institute of Linguistics]: <http://jazykovaporadna.sme.sk/>

Ministerstvo kultúry SR – Kodifikačné príručky [Ministry of Culture of the Slovak Republic: Codification manuals]: <http://www.culture.gov.sk/posobnost-ministerstva/statny-jazyk/kodifikacne-prirucky-c6.html>

Omnia Slovaca III Maior (18.01) (the in-house corpus of the Ľ. Štúr Institute of Linguistics): http://sketch.juls.savba.sk/bonito2/run.cgi/first_form

Ortograficko-gramatický slovník [Orthographic and Grammatical Dictionary]: <http://lex.juls.savba.sk/>

Slovenský národný korpus – prim-6.1-public-sane [Slovak National Corpus – prim-6.1-public-sane]. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2013: <http://korpus.juls.savba.sk>

Milujeme slovenčinu [We love Slovak] (the website of amateur linguistic advisors): <http://www.milujemeslovenčinu.sk/>

Dictionaries

Jarošová, Alexandra (ed.). 2015. *Slovník súčasného slovenského jazyka*. [Dictionary of the Contemporary Slovak Language.] M – N. (3rd volume.) Bratislava: Veda.

Káčala, Ján & Pisárčiková, Mária & Považaj, Matej (eds.). 2003. *Krátky slovník slovenského jazyka*. [Concise Dictionary of the Slovak Language.] 4th edition. Bratislava: Veda.

Kráľ, Ābel. 2009. *Pravidlá slovenskej výslovnosti*. [Slovak Pronunciation Rules.] 2nd edition. Martin: Matica slovenská.

Považaj, Matej (ed.). 1991. *Pravidlá slovenského pravopisu*. [Rules of Slovak Spelling.] Bratislava: Veda.

Sokolová, Miloslava & Jarošová, Alexandra (eds.). 2016. *Ortograficko-gramatický slovník*. [Orthographical and Grammatical Dictionary.] (online). (<http://lex.juls.savba.sk/>)

Acknowledgement

The research reported in this paper has been partially supported by the Slovak Science Foundation (VEGA), grant No. 2/0017/17 *The Dictionary of Contemporary Slovak Language – 6th Stage*.

Alexandra Jarošová
Jazykovedný ústav L. Štúra Slovenskej akadémie vied
Panská 26
811 01 Bratislava
Slovenská republika
alexandra.jarosova@juls.savba.sk

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 3[cit. 2019-11-30]. Available on web page http://www.skase.sk/Volumes/JTL41/pdf_doc/01.pdf. ISSN 1336-782X