The Dictionary of the Contemporary Slovak Language: A Product of Tradition and Innovation

Alexandra Jarošová & Vladimír Benko

Abstract

After having published the first two volumes of a multivolume monolingual dictionary of Modern Slovak, herein we try to summarise the basic concepts as they have been implemented in the actual dictionary text and introduce some extralinguistic and linguistic contexts relevant to our language and political situation. Both the traditions and innovations that have influenced the actual lexicographic decisions are presented. The extralinguistic contexts are represented above all by the existence of a special linguistic institution authorised to issue codification publications, as well as by the existence of the 'Act on the State Language', the amendment to which was passed in 2010. In Slovak lexicography, the linguistic contexts are governed by two contradicting traditions of prescriptivism and descriptivism. The presented discussion of the macro- and microstructure of the dictionary introduces some novel lexicographical solutions.

1. Introduction

The new monolingual explanatory and exemplification dictionary *Slovník súčasného slovenského jazyka* (*SSSJ*, The Dictionary of the Contemporary Slovak Language) has been envisaged as a multivolume lexicographical work of an *academic type*. So far two of its planned eight volumes have been published (Buzássyová and Jarošová 2006, Jarošová and Buzássyová 2011). The dictionary is the product of the basic research of the Ľudovít Štúr Institute of Linguistics of the Slovak Academy of Sciences. In addition to fulfilling the role of offering the principal lexicographical description of the Slovak national language, the dictionary (by including a highly extensive number of grammatical data) also serves as a kind of substitute for grammatical research (the only academic morphology of Slovak was published in 1966, and no academic syntax has yet been appeared). The dictionary describes the parameters of the lexical items characterised by their 'nation-wide' occurrence. Hence, it is not intended to be just a dictionary of *literary* language, as it also includes lexical items frequently used in informal communication. The *SSSJ* covers the lexis from about the 1940s onwards, which is the generally accepted boundary of the contemporary forms of Slovak.

2. Extra-linguistic contexts

2.1. *The scientific-organisational factors*

Of direct importance for the dictionary is the existence of a specialised institution whose mission is to produce codification publications. Such a status was acquired by our Linguistics Institute in 1949. This situation, which is paralleled in several other countries, socio-linguistically contrasts with the situation where the authority of a dictionary is not supported by any state institution, but rather by a renowned publishing house. Thanks to supplementary financial support provided by the Ministry of Education and the Ministry of Culture, the *SSSJ* has a sufficiently large team (thirteen authors, three editors, one computer linguist and sixty seven terminological consultants).

2.2. *The linguistic-political factor*

The status of the *SSSJ* is also influenced by the existence of the State Language Act of 1995, with its newer wording of 2010. This Act regulates the usage of Slovak as the state language (i.e. national language) in public communication, which is not an unusual situation as such. The problematic aspect of the Act is the fact that the state language is defined as being codified, and the Ministry of Culture (through the agency of the Central Linguistic Council as an advisory board to the minister) has a voice of authority with regard to the form of the codified state language. In the State Language Act, it is stipulated that 'no such interference into the codified form of the state language which is in discrepancy with its principles is admissible'. Thus, the Ministry has the authority to decide which changes suggested by experts are in discrepancy with its principles and which will not be included into the codification publication. On its Internet page, the Ministry presents a list of publications, including dictionaries, which have the status of codification publications. Due to the fact that the *SSSJ* has described some linguistic facts in compliance with real linguistic norms but in discrepancy with the previous codifications, and did not submit these changes to the Ministry of Culture for approval, it has not been granted a note of approval of its status as a codification publication. Another publication traditionally produced by our Institute – *Pravidlá slovenského pravopisu* (The Rules of Slovak Orthography), part of which is a dictionary – is routinely submitted for approval to one of the governmental institutions as it is a compulsory constituent of the school curricula. In the case of dictionaries, the requirement of such an approval process seems problematic to many linguists because of the fact that it instigates the following dogmatic interpretation of standard Slovak: (a) the standard (literary) norm includes only the linguistic phenomena delimited by the codification publication regardless of their actual usage; and (b) within monitoring the observance of the standard language, the particular text is mechanically compared to the given manual. Such an approach can hamper linguistic creativity; it can have its impact upon the natural transition of lexical items from one stylistic register to another, and ultimately it can also deform the scientific research. Within the existing status quo, the Ministry of Culture does not publicly criticise the dictionary, and the authors of the *SSSJ* are convinced that a large majority of the changes that the *SSSJ* introduces will become part of the updated codification publications.

3. Linguistic contexts

An official codification publication, i.e. one published by or approved by a governmental institution, could only be published after 1918, i.e. after the establishment of the first Czechoslovak Republic. In the period between the world wars (1918 – 1939) no official explanatory dictionary was published. In the given period, the work concentrated on formulating the orthographic rules and accompanying them with a dictionary. Within nine years, two official editions of orthographic works were published. They represented two differing ideological attitudes and codified different sets of linguistic phenomena. The first publication entitled *Pravidlá slovenského pravopisu* (The Rules of Slovak Orthography; Vážný 1931) was based on the constitutional construct of the first Czechoslovak Republic: i.e. the proposition of the existence of a so-called 'Czechoslovak language' in its Slovak version. Part of the Slovak public reacted most critically and disapprovingly to these *Rules*. On the other hand, the second publication, *Pravidlá slovenského pravopisu* (The Rules of Slovak Orthography; Baník, 1940), was based on a certain idealised concept of the language and presented Slovak as a structurally highly regular and very self-contained language that was to be cleaned of any foreign, non-systemic elements (i.e. above all from Bohemisms). The

Slovak public witnessed the birth of the first modern explanatory dictionary as late as in 1959, when the first volume of *Slovník slovenského jazyka* (A Dictionary of the Slovak Language; Peciar 1959 – 1965) was published. Many criticised the dictionary as only a mere registration of usage covering a period of 100 years and as not having any normative character (on the basis of excerpts it failed to exclude from the standard numerous supposed or real Bohemisms, while on the contrary it excluded dialecticisms). This critical position found its reflection in the conception of *Krátky slovník slovenského jazyka* (A Concise Dictionary of the Slovak Language; Kačala and Pisárčiková 1987). This was based on an envisaged ideal of the language which was to be respected within standardising the linguistic usage. In its 'purificational-constructional' spirit, the dictionary 'purified' the lexis from unnecessary units (mostly Bohemisms) and constructed linguistic elements complying with the system. The *SSSJ* as a dictionary chronologically following after *Krátky slovník slovenského jazyka* is thus being created in an environment of a certain codification continuum characterised by the alternative application of contradictory codification attitudes.

4. The descriptive and prescriptive dimension of the SSSJ

What kind of language is the *SSSJ* trying to describe? In short, we can say that it is the 'real' language (in this primary approximation, without the attribute 'literary') as manifested in actual language usage. We have been working with the notion of a 'real norm', which in contrast to usage is on a higher level of abstraction; it is *langue*-oriented and is more compatible with the character of lexicographic reflection. Within the attempts at establishing the norm, a decisive role is played by the frequency manifested in large databases, although we cannot count on the absolute objectivity of the description in consequence of the qualitative limits of the corpora. Any further splitting of the norm into a standard one and any other one (i.e. non-standard or non-classifiable) is the result of imposing artificial linguistic criteria (cf. Nebeská 1999, Cvrček 2008). As language does not just represent a mere instrumentality, and as Slovaks perceive their language as having a highly positive value, the legitimacy of this concept makes it acceptable and applicable to our dictionary. We tend to avoid using the label 'incorrect'. The non-standard character of a lexical unit, i.e. any restrictions of its usage only to the sphere of the informal, is indicated by the labels 'slang' and 'substandard' (loan words, mostly non-adapted Bohemisms bearing the structural features of the source language). The *SSSJ* reflects the development of linguistic standards to a larger extent than this is done and can be done by a concise dictionary. This is manifested for instance by presenting a larger number of lexical, pronunciation and morphological variants, as well as by presenting a different normative-stylistic evaluation of some words in comparison to their actual codification. In the *SSSJ* the identified lexical and morphological variants are ordered according to their frequency.

5. The manner of presentation: selected aspects

5.1. *Macrostructure*

The main novelty within the dictionary macrostructure is the *'autonomisation' of the list of headwords*. Above all, this concerns the separate (i.e. autonomous) listing of derivatives: presented in a new manner in the dictionary are the semantic descriptions of denominal relational adjectives (*golier* n. 'collar'→ *golierový* adj. 'pertaining to collar', *baník* n. 'miner'→ *banícky* adj. 'pertaining to mines, miners, or mining'), deadjectival adverbs (*jemný*

adj. 'gentle'→ *jemne* adv. 'gently'), and deadjectival nouns (above all, abstract nouns: *jemný* 'gentle' → *jemnosť* 'gentleness'). While in previous dictionaries these lexical groups were presented as run-ons attached to the headwords of their motivating noun, in the *SSSJ* they are presented separately.

The presentation of lexical units within headwords and sub-headwords (i.e. of the lexical units having an explanation of their own) has influenced the *manner of the presentation of multiword lexical units*. The main criterion for classifying multi-word lexical units is constituted by the type of their nominative function: (a) phrasemes or phraseological units – lexical syntagmas or sentences, the nominative function of which specialises primarily in conveying the evaluative and expressive attitudes of the speaker (i.e. their semantic content comprises evaluative and expressive elements); (b) lexicalised word combinations – adjective-noun compounds, verb-noun constructions with an abstract noun; and (c) multiword (technical) terms – lexical units having an exact definition of meaning. How is this classification related to the idiomaticity of word combinations and how do we understand this idiomaticity? The idiomacity of a word combination is based on the anomalous (irregular) combinatorics of their elements, 'of which at least one does not function in the same way in another combination' (Čermák 2007: 147). Idiomacity understood in this way is present in all phrasemes, and it is also possible (though in Slovak not prototypical) in lexicalised word combinations and multiword terms. Hence, the main classification feature for separating phrasemes is not idiomaticity but their pragmatic meaning. Phrasemes are placed at the end of the respective entries (in the case of a polysemic entry after its last sense) and their status is similar to that of headwords (they are indicated in colour and accompanied by explanations). Multiword terms, for example the hydrographic term *plavebná hĺbka* 'navigable depth' or the term in photography *hĺbka ostrosti* 'depth of focus', are included in the first sense of the primarily non-terminological headword **hĺbka** 'depth', but they are separated from the ordinary word-combination data by a special symbol. Similarly, lexicalised word combinations, e.g. *rybia kosť* 'herringbone, fishbone' (a common denotation of the textile-related term 'broken twill weave') and *špiková kosť* 'marrowbone', are separated from 'free-word combinations' (e.g. *hovädzia kosť* 'beef bone') in the second sense of the entry **kosť** 'bone'.

5.2. *Microstructure*

In constructing a definition we take into consideration the model of the meaning of the word as a potential complex of knowledge about the denoted entity or situation, as well as the complex of evaluations, attitudes and connotations present in the consciousness of an average member of the language community. This perception of meaning (influenced by cognitive linguistics) implies a diversion from the strictly structural conception of meaning as presented in previous lexicographical works. Instead of binary oppositions, we prefer a scale-type presentation of the lexical-semantic phenomena. These theoretical points of departure are reflected in the structure and information content of the entry by means of its *contextualisation* (i.e. enrichment of the definitions by elements of typical contexts using a wide range of examples) and by means of its *pragmatisation* (using supplementary and/or explanatory notes of a pragmatic nature).

6. Sources of lexical evidence and computational support

From the very beginning of the *SSSJ Project*, language corpora use had been considered as an important source of lexical evidence. Though the technical infrastructure (*WordCruncher* under MS-DOS on a couple of PCs) and the size of the corpora (20 million tokens) in the early 1990s may make us smile from today's perspective, the situation has kept improving each year and after establishing the *Slovak National Corpus* Department in our Institute in 2002, the size of the corpora started to grow (almost) exponentially. At present, our lexicographers can use two large Slovak corpora, the referential *Slovak National Corpus* (Version 5.0, 700+ million tokens) and a web-based *Bruna Corpus* (created at Masaryk University, Brno, Czech Republic, 850+ million tokens). To access both corpora we can use the *Sketch Engine* (Kilgarriff et al. 2004) installed on our servers using specialised sketch grammar optimised for lexicographic usage and preferring recall over precision so that no important collocation information is omitted (Benko 2010).

References

A. Dictionaries

Baník, A. A. (ed.) 1940. *Pravidlá slovenského pravopisu s pravopisným.* Turčiansky Sv. Martin: Matica slovenská.

Buzássyová, K. and A. Jarošová (eds.) 2006. *Slovník súčasného slovenského jazyka.* (1. zv.) A – G. Bratislava: Veda.

Jarošová, A. and K. Buzássyová (eds.) 2011. *Slovník súčasného slovenského jazyka.* (2. zv.) H – L. Bratislava: Veda.

Kačala, J. and M. Pisárčiková (eds.) 1987. *Krátky slovník slovenského jazyka.* (First edition.) Bratislava: Veda.

Peciar, Š. (ed.) 1959–1965. *Slovník slovenského jazyka. I. – V. diel.* Bratislava: Vydavateľstvo Slovenskej akadémie vied.

Vážný, V. (ed.) 1931. *Pravidlá slovenského pravopisu s abecedným a pravopisným slovníkom.* Turčiansky Sv. Martin: Matica slovenská a Praha: Štátne nakladateľstvo.

B. Other Literature

Benko, V. 2010. 'Optimizing word sketches for a large-scale lexicographic project.' In SKEW Workshop. Ljubljana. Available at: http://videolectures.net/korpusi2010_benko_ows/

Cvrček, V. 2008. *Regulace jazyka a koncept minimální intervence.* Praha: Nakladatelství Lidové noviny – Ústav českého národního korpusu (Institute of the Czech National Corpus).

Čermák, F. 2007. *Frazeologie a idiomatika česká a obecná.* [parallel title] *Czech and General Phraseology.* (First edition.) Praha: Nakladatelství Karolinum.

Kilgarriff, A. et al. 2004. 'The Sketch Engine.' In G. Williams and S.Vessier (eds.), *Proceedings of the eleventh EURALEX International Congress EURALEX 2004 Lorient, France, July 6-10, 2004.* Lorient: Université de Bretagne-Sud, 105–116.

Nebeská, I. 1999. *Jazyk – norma – spisovnost* (Language – Norm – Standard Character.) (First edition.) Praha: Nakladatelství Karolinum.