Terminology and Terminological Activities in the Present-Day Slovakia

Jana Levická

E. Štúr Institute of Linguistics Slovak Academy of Sciences Bratislava, Slovakia janal@korpus.juls.savba.sk https://data.juls.savba.sk/std

Abstract. The author of this paper aims to present current situation and problems of terminology monitoring and administration in Slovakia that lacks appropriate terminology discussion and terminography activities. In order to raise the awareness of the terminological issues, a centralised national term base was created intending to draw on the subcorpora of respective fields. The subcorpora not having been completed so far, terminology entries are to be filled in with available internet data. The author draws special attention to the key information of the terminology entry – definition, which, however, is not always to be found on the internet. Therefore, the team opted for the so-called defining context as a temporary substitute. The structure and content of both definition and defining context is to be analysed, which may serve for further semi-automatic retrieval in the specialised corpora.

Slovak language has seen spontaneous and uncontrolled accumulation of terms for more than 17 years due to the political, economic and social transformation of Slovakia after 1989, which resulted in coinage of excessive and often unnecessary terminological variants of both domestic and foreign provenience.

In spite of a rich history of terminological activities, Slovak society has been facing a double vaccum in its post-war history – on one hand in terms of analysis and development of foreign and domestic terminological theories, technologies and methodologies and on the other studies comparing foreign and Slovak terminological systems. Moreover, the new political and economic situation has caused a massive braindrain in the academic and scientific sphere, thus Slovakia lacks available full-time and skilled terminologists not to mention sufficient linguistic curricula and existent terminological education.

Although terminological activity in Slovakia has suffered the greatest fallout in the last 50 years, it was not suppressed completely. Quality bilingual or unilingual specialised dictionaries have been sporadically published, quality theses elaborated and articles on the theory of terminology written and published especially in the revue $Kultúra\ slova$ but have only seen modest feedback.

Some institutions, aware of the urgent need, started with scarce but key terminological activities, i.e. setting up of terminology databases – e.g. the National Bank of Slovakia and the Slovak Institute of Technical Normalisation. However, those are only domain limited databases with specific, very narrowly defined aims – the former is an in-house tool for employees of the bank while the latter is used by the creators and translators of technical norms. As a mater of fact, these banks do not provide any access for lay public.

At the same time, there has been an external demand for consistent and unified Slovak terminologies for the purpose of drafting and translating European legislation into Slovak as they have been revealed to be unsatisfactory.

1 Place of a corpus in the context of terminological activities

It is common knowledge that effective specialised communication requires unambiguous terminology. The contemporary Slovak state of affairs does not contribute to coherent and intelligible science neither for specialists, producers and lawmakers nor for teachers, translators and interpreters. Therefore terminology monitoring, coordination, analysis of the special vocabularies and unification of pertinent results in the form of glossaries, dictionaries, terminological standards or terminology databases, which present-day Slovakia lacks, is considered by the L. Štúr Institute of Linguistics of the Slovak Academy of Sciences to be a priority.

Centralisation of various terminologies under one administration, continuous modification and updating of term records and narrow collaboration of terminological boards, translators and specialists is nowadays regarded to be the only way of terminological harmonisation, and consequently standardisation. Contemporary terminological tendencies stress the model of the text and corpus approach as a sine qua non prerequisite of every terminological project. The process of systematic gathering of terms is based exclusively on representative corpora, supervised and validated by specialists and terminologists. As Sager (1990:131) puts it, information extracted from a text represents a reliable indicator of changes and ensures the only plausible data for building and revising terminological records.

It is therefore only natural that a terminology database project be started by the Corpus Department of the Institute for it had all the resources and tools at its disposal – textual base of the Slovak National Corpus itself and software ones for automatic annotation of Slovak texts.

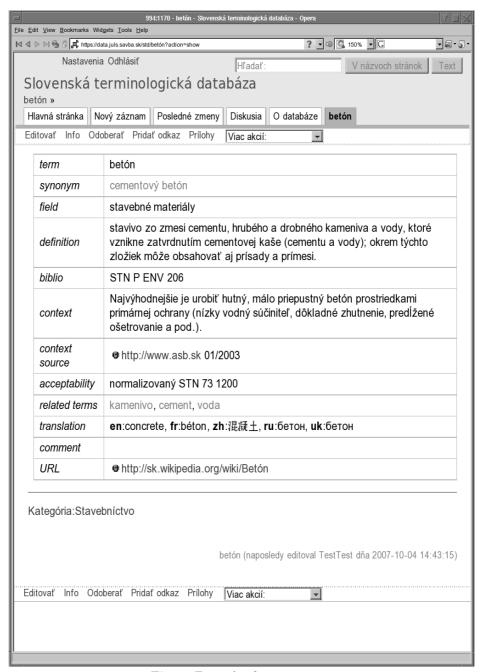
2 Project of the Slovak National Terminology Database

The SNC project aims to set up a terminological database provided with both conceptual and linguistic information, inspired by foreign examples, mostly Canadians but, of course, adapted to Slovak needs and present-day possibilities of the E. Štúr Institute of Linguistics. The team expects to cooperate and exchange the data with leading European database IATE which is why the EUROVOC 4.2 Thesaurus was chosen as the classification system.

The starting point of the Slovak Terminology Database Project dates back to the autumn of 2005 when the SNC team 2005 launched the analysis of existing terminology databases, subsequently proceeding to the design of the term record layout based partly on translation and interpreting needs survey. The team chose and adapted the appropriate software for the database along the way. The SNC policy of text acquisition also had to be modified and the focus was shifted towards economic and legal texts for the purpose of creating specialised subcorpora and further automatic extraction of terms and possibly definitions as well as other terminological data from specialised corpora.

Project methodology, as we already mentioned, has drawn inspiration from the textual terminology approach to the terminology extraction of lexical units – potential terminological units from running specialised texts and identification of the concept they refer to. No less importance is to be attached to the generation of software tools for automatic extraction of terms and possibly definitions from specialised subcorpora.

As far as the data categories of the term record are concerned, the team opted for an 11-item term record containing 7 obligatory fields. In order to satisfy the needs of professionals, lay public and last but not least the translation and interpreting public, the team decided to not only make the field of definition and domain obligatory, but also context, related terms and sources of definition and context. The remaining 4 optional fields feature synonym, foreign language equivalent, comment and links to reliable web pages.



 ${\bf Fig.~1.}$ Example of a term entry.

3 Changes in the project strategy

However, the project did not receive funding from the state run Slovak Research and Development Agency. The team had to take a different path and began to work within a reduced scale, i.e. instead of creating new records by filling in obligatory fields (1, 3, 4, 5, and 6), the emphasis was shifted on reusing and adapting existing quality terminology resources published in Kultúra slova in particular, as the team received a copyright license for their non-commercial use as well as some of those that had been elaborated by our close collaborators.

For the time being, the database offers almost 3000 terminological records covered by 8 domains (Astronomy, Security and Law, Migration Policy, Construction, Corpus Linguistics, Phraseology, Phonetics and Phonology, Bilingualism, Civil Security, Historical Linguistics, Fire Protection) with more or less completed terminology fields – term and usually definition, source, less often synonym, sometimes related terms, and comment.

In order to meet one common form of term record we had to proceed to the harmonisation of different editorial practices. This alternative thus brought about different issues to discuss and deal with.

On the formal level, the most frequent flaws include term records written in incorrect Slovak, excessive punctuation within the definition and formal treatment of polysemous terms.

As far as the content of the term record is concerned, problem issues cover harmonisation of EUROVOC descriptors with the classification of the original terminological ones, which are usually fine-grained; evaluation of relevancy of terms belonging to the terminology of a specific author or school; identification and distinction of different types of related terms; ascribing the status of terms; limitation of inconsistency of terms records as such; treatment of nomenclatures and accompanying data etc.

In spite of all rules and efforts, definition-related discrepancies include incompleteness, inconsistency, amateurishness and subjectivity as well as distinction and splitting of original entries into definition, context and comment fields.

4 Definition and context analysis

The second part of the paper will be focused on the key obligatory field of the proposed term record – definition. Upon a brief introduction dealing with its typology and function, the attention will be paid to its delimitation and

structure, which will subsequently help to draw analogy with the so-called defining context as the nearest substitute of the definition.

4.1 Definition rôle and typology

Definition represents a sort of microsystem consisting of hierarchically ordered characteristics of a concept and their relations, which enable to describe, circumscribe and distinguish the concept. However, features included in a definition reflect the concept structure but can never cover the totality of a concept (Seppälä 2004: 37), hence the origin of variant definitions of the same concept.

Terminological practice has recorded numerous typologies of definitions based on different perspectives – e.g. situation of use, defining mode, formal composition, content of the definition, rôle, and editing practice, the choice being dictated by the target audience, aim of terminographic project and respective domain.

Terminological theory fosters traditional and most frequent Aristotelian definition, which begins with the nearest superordinate concept and specific features (genus term and differentia or characteristics), i.e. "systematically identifies a concept with respect to all others in the particular subject field" Sager (1990:42). This so-called ideal definition with specific editing criteria to follow is referred to as classic, intensional or comprehensive definition.

Example: Samozhutniteľný betón: betón, ktorý je schopný tiecť a spevňovať sa účinkom vlastnej hmotnosti, úplne vyplniť debnenie, aj vo vysoko vystuženom priereze, za súčasného udržiavania homogenity a bez potreby hocijakého dodatočného zhutnenia.

"Samozhutniteľný betón" is defined with the aid of the closest genus, which is "concrete" as well as the characteristics, which distinguish this specific concrete from all other types of "concretes":

- schopný tiecť a spevňovať sa účinkom vlastnej hmotnosti,
- úplne vyplniť debnenie, aj vo vysoko vystuženom priereze, za súčasného udržiavania homogenity a bez potreby hocijakého dodatočného zhutnenia.

Comprehensive definition has its counterpoint in the **extensional definition** that ISO 740 defines as "an enumeration of all species which are all on the same level of abstraction".

Example from the STD: materiálno-technické vybavenie jednotiek požiarnej ochrany: vybavenie jednotiek požiarnej ochrany zahŕňajúce požiarnu techniku, vecné prostriedky požiarnej ochrany na výkon odborných služieb a špecializovaných činností, hasiace látky, ako aj ochranné prostriedky na účinné vykonávanie zásahu a činnosti na požiarnej stanici.

It is a common phenomenon to find so-called mixed definitions that consist of a comprehensive as well as extensional part.

4.2 Context rôle and typology

What is meant by the "context"? **Context** as such is defined in the ISO 12620 as a "text which illustrates a concept or the use of a designation", or "a text or part of a text in which a term occurs". The two definitions indicate several functions of the context exploited in terminological works according to which it is possible to distinguish three types of contexts:

- 1. **language context** that can be identified with *lexicographic example*, which is a simple occurrence of a term indicating neither conceptual nor linguistic information.
- 2. **linguistic context** is a context that illustrates the linguistic function of a term in discourse but provides no conceptual information e.g. typical syntactic structures or collocations.
- 3. **defining context** that Canadian term base Termium considers as the one that sheds light on the distinctive characteristics of the concept. ISO 12320:1995 informs that it "contains substantial information about a concept but does not possess the formal rigor of a definition", i.e. the substantial information can cover essential characteristics of the concept, its purpose, consequence of the action/event etc. Sue Ellen Wright offers a slightly extended ISO definition and at the same time a less restricted one: "defining context contains definitive information that may look very much like a definition, but is *incomplete* or doesn't have the right form for a definition".

In conclusion, the comprehensive definition is made up of a genus term and essential characteristics while the defining context does not have to explicitly express the superordinate genus term and all the essential characteristics but must include enough information to identify the concept.

Due to the lack of definitions, not speaking of the quality of existing ones, it is planned to use the defining context in Sue Ellen Wright's sense, as a provisional terminological data until a proper definition is found or formulated as it is the closest one to the definition and the most pertinent one for classification and identification purposes of a term.

5 Definition and context mining: Slovak National Corpus vs Internet

Maintaining the idea of creating new term records in the near future with only a limited team of unskilled persons (for the specialists can be contacted only for the revision/validation process), the team has decided to use ready-made definitions and defining contexts that may be available in our Slovak National Corpus (SNC) and Internet.

Therefore, research focused on types, frequency and quality of definitions and defining contexts in both sources has been carried out within domain specific areas. For the purpose pilot research on the prescriptive domain of construction was selected because, i.e. terms referring to construction materials. The idea was to identify the linguistic and conceptual structure of the definition and thus find a repeating structure or pattern of defining elements, i.e. genus term and differentia and key words expressing them might enable semi-automated extraction of both terminological types of data.

The analysis is based on the French-written thesis by Selja Seppälä (2004) titled *Conceptual Composition and Formalisation of the Terminograhical Definition* and her two typologies: 13-item typology of genus conceptual classes and 22-item typology of differentia, which she used for manual annotation of a corpus made of 500 definitions after their identification and isolation.

5.1 Corpus search

Since a construction subcorpus is not available yet, we started our analysis by searching the 3 juls-all version of the SNC within the text annotated as TEC domain, which yielded 801 occurrences of the lemma $bet \delta n$ by means of regular expressions. On the basis of the collocation statistics we could identify the most frequent complex terms out of more than 40.

As for the typology of sources, the virtual subcorpus consists of specialised magazines ASB and $Materi\'{a}lov\'{e}$ $in\'{z}inierstvo$ (25 issues and 2 issues respectively), a commercial leaflet and a semi-specialised book representing the two remaining sources. Occurrences of $bet\'{o}n$ from a popular internet magazine inZIne and the specialised IT magazine PC REVUE had to be classified as unacceptable and therefore left out from further analysis.

According to the Canadian *Handbook of Terminology* classification of relevant and reliable terminological sources, specialised and popularised periodicals are ranked 4th while the brochures and publicity flyers are 5th.

However, manual search of the abovementioned occurrences revealed to be highly disappointing for it yielded only 3 relevant results: 1 definition and 2 defining contexts deriving from the same source – ASB magazine. We present their annotated structure and content as follows:

Definition

SSC je [betón] inanimate—artificial [s veľkou pohyblivosťou a schopnosťou tiecť bez pôsobenia vonkajších dynamických síl] physical property, [s mimoriadnou odolnosťou proti rozmiešavaniu a segregácií hrubých zložiek čerstvého betónu] utility. Source: ASB -2004/07

Defining contexts

Výroba sa sústreďuje do centrálnych výrobní betónu, kde sa vyrába tzv. transportbetón, t.j. čerstvý betón inanimate→artificial [mäkkej alebo tekutej konzistencie] physical property, ktorý [sa prevezie auto-

domiešavačom/ INSTRUMENT a ktorý [sa na stavbe ukladá do debnenia/ utility. Source: ASB 2003/07

Vysokopevnostný betón HSC (High Strength Concrete) [patrí do pomerne nedávno (v roku 1993) vytvorenej skupiny vysokohodnotných betónov – HPC] WHOLE (High Performance Concrete), pre ktoré je charakteristická [pevnosť v tlaku vyššia než 65 MPa] measurable property Source: ASB 2005/07-08

Comparing the two defining contexts to the definition, the first one introduces defining elements by means of the explanatory structure t.j. (i.e.), which is synonymous to the verb be. The other uses a sort of paraphrase of the conceptual class indicating membership of this concrete to a specific group and differentia are launched with the syntactic structure: je $charakteristick\acute{y}$.

5.2 Internet search

The research proceeded by searching definitions and defining contexts of 12 complex terms in their nominative form selected with respect to their highest absolute frequency in the 3-juls-all SNC within texts annotated as TEC domain.

Internet sources include again the specialised magazine ASB, two association portals (http://www.betonracio.sk and http://www.beton.sk), two educational pages, European Directive on Self-Compacting Concrete and the individual professional portal http://www.dalnice.com.

	Complex term	Frequency in the SNC	Google search occurrences
1.	čerstvý betón	72	153
2.	pohľadový betón	31	92
3.	podkladový betón	25	116
4.	asfaltový betón	19	82
5.	cementový betón	11	32
6.	predpätý betón	9	106
7.	vysokopevnostný betón	8	20
8.	vystužený betón	7	40
9.	samozhutňujúci betón	6	14
10.	vodotesný betón	5	51
11.	vysokohodnotný betón	5	23
12.	samozhutniteľný betón	4	21

Table 1.

In spite of a relatively high number of occurrences, this research resulted in identifying only 5 defining contexts and 6 definitions of googled complex terms. In the case of five complex terms we did not get any relevant result. We present the annotated conceptual structure of genus terms and differentia as follows:

Definitions

- 1. **čerstvý betón** betón inanimate→artificial, ktorý je [úplne zamiešaný a je ešte v takom stave, ktorý umožňuje jeho zhutnenie zvoleným spôsobom] condition. http://www.beton.sk
- vyysokopevnostný betón betón inanimate→artificial, ktorý má [pevnostnú triedu v tlaku väčšiu ako C 50/60 (B60)] measurable property [pre obyčajný a ťažký betón] type a [LC 50/55 (B55)] measurable property [pre ľahký betón] type. http://ww.beton.sk.
- 3. samozhutniteľný betón je inovovaný betón INANIMATE→ARTIFICIAL, ktorý [nevyžaduje vibrovanie pri ukladaní a zhutňovaní.] CONDITION http://w.savt.sk/dokumenty/public/eur-pske-smernice-presamozhutnite-318-ny-bet-n/view.html
- 4. samozhutniteľný betón betón inanimate→artificial, ktorý je [schopný tiecť a spevňovať sa účinkom vlastnej hmotnosti] physical property, [úplne vyplniť debnenie, aj vo vysoko vystuženom priereze, za súčasného udržiavania homogenity a bez potreby hocijakého dodatočného zhutnenia] utility
 - http://ww.savt.sk/dokumenty/public/eur-pske-smernice-pre-samozhutnite-318-ny-bet-n/view.html
- 5. **cementový betón** *je zmes* inanimate—artificial *vysoko kvalitných drtených kamenív, cementu a vody* content.
 - http://www.dalnice.com/pojmy/slovnicek.htm
- 6. **asfaltový betón** − *je to [zmes]* INANIMATE→ARTIFICIAL [hutného kameniva [s uzavretou zrnitosťou zahorúca obalená asfaltom] PHYSICAL PROPERTY] CONTENT. http://www.asb.sk

All definitions, being comprehensive ones, rank defined terms among the conceptual classes inanimate—artificial expressed by the designation /bet'on/ linked with the rest of defining elements by verb junctors be or have, which confirms Seppälä's findings (2004:138).

As for the categories of characteristics, the most frequently identified are measurable property (2x) and condition (2x), utility, type, content and physical property.

Defining contexts

- 1. **čerstvý betón** je [v nezatvrdnutej forme dopravovaný na miesto spotreby] condition [stavenisko] place indicator, kde je [pripravený pre uloženie do debnenia a po zatvrdnutí je spravidla hlavnou nosnou časťou stavebných konštrukcií] utility
 - http://www.savt.sk/r-zne/charakteristika-vyroby-transportbet-nu.html
- 2. samozhutniteľný betón možno charakterizovať ako [extrémne tekutý] physical property [betón] inanimate—artificial, ktorý [dokáže úplne vyplniť priestor debnenia alebo formy a zhutniť sa bez použitia vibrácie alebo iného spôsobu zhutňovania]. Utility
 - http://www.betonracio.sk
- 3. Hlavnou charakteristickou vlastnosťou **SCC** je [schopnosť tečenia čerstvého betónu bez pôsobenia vonkajších dynamických síl] physical property, [odolnosť proti rozmiešaniu a segregácii hrubých zín kameniva a schopnosť zhutnenia vlastnou hmotnosťou] utility.
 - http://www.asb.sk
- 4. **Predpätý betón** *vzniká* [kombináciou betónu a predpätej výstuže] CONTENT http://www.soustavebne.sk/betony/predpaty_beton.htm
- 5. **Vodotesný betón** *je [betón]* inanimate—artificial, *ktorý [odoláva tlakovej vode tak, že na jeho vzdušnej strane nevzniknú viditeľné priesaky, prípadne vlhké škvrny]* physical property.

http://fzki.uniag.sk/

Only two out of five defining contexts rank defined terms among conceptual classes inanimate—artificial expressed by the designation $/bet\acute{o}n/$ linked with the rest of defining elements by verb junctors be, vznikat or by means of classifying paraphrases $mo\check{z}no$ charakterizovat, charakteristickou vlastnostou. Lack of formal rigour is evident in the wording of all texts.

As for the categories of characteristics, we identified most frequently physical property (3x) and utility (3x), followed by condition, place indicator, content.

Conclusion

Due to the insignificant number of found definitions and defining contexts, their comparison is not of much relevancy, moreover, only in the case of two complex terms the search revealed both defining contexts and definitions. However, analysed definitions and contexts show the same conceptual class of the genus term and three same characteristics (UTILITY, PHYSICAL PROPERTY,

CONDITION). Excerpted definitions seem to be more precise (see number of MEASURABLE PROPERTY) while defining contexts appear to be more function-oriented (see utility). The author is aware that these tendencies must be verified in a representative and balanced specialised subcorpus in order to ensure automated research of collocated defined words with their genus and differentia terms.

References

BÉJOINT, Henri: La définition en terminographie. In: Aspects du vocabulaire. Ed. P. J. L. Arnaud et Ph. Thoiron. Travaux du CRTT: Lyon, Presses Universitaires de Lyon, pp. 19-25.

BOURIGAULT, Didier, AUSSENAC-GILLES, Nathalie, CHARLET, Jean: Construction de ressources terminologiques ou ontologiques à partir de textes: un cadre unificateur pour trois études de cas. In *Revue d'Intelligence Artificielle (RIA)*, Techniques Informatiques et structuration de terminologiques, Pierrel J.-M. Et Slodzian M. (Ed.), Paris: Herměs. 2004, vol. 18, n. 1, pp. 87 – 110.

CABRÉ, Maria Teresa: La terminologie – théorie, méthode et applications. Ottawa: Armand Colin/PUO 1998.

ČERMÁK, F. et alii.: Manuál lexikografie. Praha: H&H 1995.

GAUDIN, François: Socioterminologie. Une approche sociolinguistique de la terminologie. Bruxelles: De Boeck et Larcier 2003.

GESCHÉ, Véronique: Évaluation des définitions d'ouvrages. In Meta, 1997, roč. 42, n. 2, pp. 374-390.

HORECKÝ, Ján: Základy slovenskej terminológie. Bratislava: Vydavateľstvo SAV 1956.

ISO 704 (2000): Travail terminologique – principes et méthodes. International Organization for Standardization.

ISO 860 (1996): Travaux terminologiques – Harmonisation des termes. International Organization for Standardization.

ISO 1087-1 (2000): Travaux terminologiques – Vocabulaire. International Organization for Standardization.

ISO 12620 (1999): Aides informatique en terminologie. International Organization for Standardization.

KOCOUREK, Rostislav: La langue française de la technique et de la science. Wiesbaden: Brandstetter 1991.

LERAT, Pierre: Les langues spécialisées. Paris: PUF 1995.

LINO, Maria Teresa Rijo da Fonseca: Base de données textuelles et terminographiques. In Meta, 1994, vol. 39, n. 4, pp. 786 – 789.

MASÁR, Ivan: Príručka slovenskej terminológie. Bratislava: VEDA 1991.

NAZARENKO, Adeline, HABERT, Benoît, SALEM, André: Les linguistiques de corpus. Paris: Armand Colin 1997.

NUOPPONEN, Anita: Terminological information and activities in the world wide web. In Christian Galinski & Klaus- Dirk Schmitz (eds.), TKE'96, Terminology and Knowledge Engineering. 92–99. Frankfurt: INDEKS-Verlag.

OTMAN, Gabriel: Les bases de connaissances terminologiques: les banques de terminologie de seconde génération. In Meta, 1997, vol. 42, n. 2, pp. 244-256.

PAVEL, Silvia, NOLET, Diane: Handbook of Terminology. Translation Bureau, 2001 http://www.translationbureau.gc.ca

SAGER, Juan Carlos: A Practical Course in Terminology Processing. Amsterdam/Philadelphia: John Benjamins 1990.

SEPPÄLÄ. Selja: Composition et formalisation conceptuelles de la définition terminographique. *Mémoire pour l'obtention du DEA*. Genève: Université de Genève 2004.

SCHWARZ, Jozef: Vybrané teoretické a metodologické problémy terminografie: poznatky z tvorby České terminologické databáze knihovnictví a informační vědy. In Národní knihovna, 2003, vol. 13, n. 1, pp. 21-41.

Internet sources

http://www.termium.com

http://www.nkp.cz/o_knihovnach/Slovnik/index.htm

http://www.yourdictionary.com

http://www.cilf.org

http://www.termisti.refer.org

http://www.erudit.org/revue/meta/

http://www.rint.org

http://www.infoterm.info/index.php

http://linux.termnet.org/ http://europa.eu/eurovoc/