

Automated Identification of Domain Preferences of Collocations

Jelena Kallas¹, Vit Suchomel², Maria Khokhlova³

¹Institute of the Estonian Language, Estonia

²Masaryk University, Czech Republic

³St. Petersburg State University, Russia

E-mail: jelena.kallas@eki.ee, xsuchom2@fi.muni.cz, m.khokhlova@spbu.ru

Abstract

This paper addresses (semi-)automatic collocations dictionary compilation in connection with the automated identification of domain preferences of collocations. The research was motivated by the process of the semi-automatic compilation of the Estonian Collocations Dictionary (ECD), where lexicographers processed a large number of terminological collocations extracted from Sketch Engine into the Dictionary Writing System EELex.

In this paper, we apply the terminology extraction module within the Corpus Query System Sketch Engine and present the results of the experiments on building military domain corpora in Russian and Estonian and extracting multiword terms. Both languages have very rich morphology and quite a large number of multiword terms, but Russian texts are well represented on the Web while Estonian ones are not. We analyze how the comparison of frequency of a collocation in a reference corpus with its frequency in a domain corpus can be used for facilitating word sketch data analysis in terms of identification of domain preference of collocations.

Keywords: collocation; multiword terms; terminological collocation; Russian; Estonian

1. Introduction

Building terminological lexicons and glossaries is a prominent task in many areas: from translators to large companies aiming to establish consistent naming in their documentation. Also for lexicographers it is quite tricky to extract terminology from texts and label it properly. As Atkins and Rundell (2008: 227) point out, domain labels play an important role in lexical databases. “A domain label indicates that *the item is used when the subject of discussion is ... (science, hockey, plumbing, poetry etc.)*”.

Traditionally, domain labels are assigned in dictionaries to word senses. However, it is also quite a common practice in collocations dictionaries. For example, the *Oxford Collocations Dictionary for Students of English* (OCDSE, 2002) presents domain specific collocations as “technical collocations” and defines them as “collocations that are used by people who specialize in a particular subject area”. Altogether, eight different subject areas are distinguished (business, computing, law, mathematics, medical, military, science and sport). In addition to these labels, more specific usage restriction, such as ‘in football’ or ‘used in journalism’, are given in brackets.

As for automated collocations dictionaries, no domain labels have been provided so far. An example of an automated collocation dictionary entry is shown in Figure 1, illustrating the lexeme “operation” in the Sketch Engine for Language Learning (SkELL) system (Baisa & Suchomel, 2014).

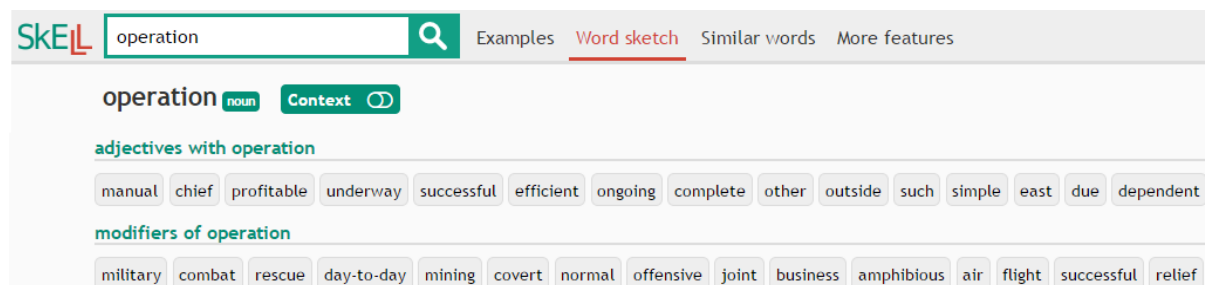


Figure 1: An example of a word sketch for “operation” in SkELL

Among collocates, there are quite a few examples of units that belong to certain domains.¹ However, there are no labels that help learners to identify whether a particular collocation is a terminological one or not.

The same problem is significant for semi-automated compilation of collocation dictionaries. A recent survey (Tiberius et al., 2015; Gantar et al., 2016) showed that acquiring lemma lists and frequency information from corpora is a common procedure, followed by the extraction of example sentences, grammatical patterns, multiword expressions, form variations and neologisms. Less frequent are automated procedures related to semantics: word senses, lexical semantic relations, definitions and knowledge-rich contexts. Authors (Gantar et al., 2016: 211) point out that when analyzing word sketch data, lexicographers still spend a significant amount of time selecting the relevant collocates and their examples under each syntactic model.

One analytical lexicographic task that is also still performed manually is the identification of terminological collocations and making decisions about whether to exclude them from the database as not relevant or to add domain labels. This process is discussed in greater detail in Section 2. This task would be made less time-consuming with the development of new approaches within corpus tools. It should be possible to automatically identify collocations that are very frequent in particular domain corpora and provide this information to lexicographers.

This idea is not a new one and it is discussed, for example, in Rundell and Kilgarriff (2001) and Rundell (2012). “Essentially it involves comparing a word’s profile in a

¹ See e.g. *military operation*, which is registered as a term in the terminology database IATE. Accessed at: www.iate.europa.eu (20 May 2017)

carefully-defined sub-corpus with its behaviour in the lexicographic corpus as a whole, in order to retrieve information about its stylistic, regional, or domain preferences” (Rundell, 2012: 28).

Figure 2 illustrates how register preference can be shown as additional information in word sketch (Kilgarriff et al., 2004) data analysis. In order to achieve it there are two subcorpora (written and spoken) compared simultaneously. The label in the upper right corner, “usually in spoken (69.9%, percentile 0.4)”, indicates that this particular word is used mostly in the spoken corpus.

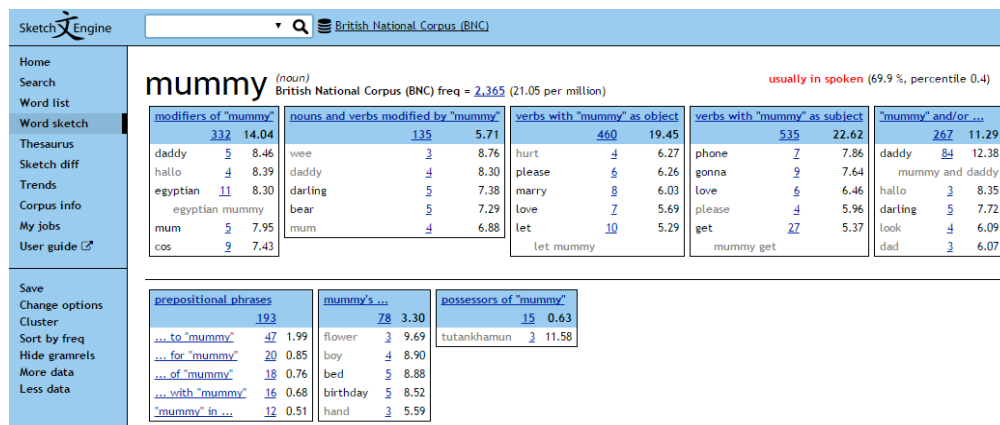


Figure 2: An example of a word sketch for “mummy” in British National Corpus, with register preference information “usually in spoken” (indicated on the right side)

Similarly, the usage of domain corpora should make it possible to apply additional filters for collocation extraction and thus to identify domain preferences of particular collocations.

In this paper, we differentiate between notions of a terminological collocation and a multiword term. For a multiword term definition, we follow the approach of Ramisch (2009). A multiword term is a term that is composed of more than one word. The unambiguous semantics of a multiword term depends on the knowledge area of the concept it describes and cannot be inferred directly from its parts (SanJuan et al., 2005; Frantzi et al., 2000). In terms of terminological collocations, we follow the conception proposed in Costa and Silva (2004). A terminological collocation can be defined as a unit consisting of a term and its collocates. For example, *баллистическая ракета* ‘ballistic missile’ can be viewed as a multiterm, whereas *запустить баллистическую ракету* ‘to launch a ballistic missile’ is a terminological collocation (however, to a certain degree the given collocation acquires the terminological status). Thus the whole item is a non-term “considering that its whole generally does not refer to a concept” (ibid). Nevertheless such terminological collocations should be presented in dictionaries with special domain labels.

2. Manual Identification of Terminological Collocations in the Estonian Collocation Dictionary Database

The Estonian Collocations Dictionary is a monolingual online scholarly dictionary aimed at learners of Estonian as a foreign or second language at the upper intermediate and advanced levels. The dictionary contains about 10,000 headwords, including single and multiword lexical items. For the automatic generation of the ECD database, the corpus query system Sketch Engine (Kilgarriff et al., 2004) functions Word List, Word Sketch and Good Dictionary Example (GDEX) were used. The main parameters used for the extraction of collocates were 1) the minimal frequency of a collocate: 10 (for the frequency I class) and five (for the frequency II class), 2) the minimal salience of a collocate: positive Dice, 3) the minimum frequency of the grammatical relation: 10, and 4) the minimum salience of the grammatical relation: positive Dice. We extracted collocates in a fixed order according to grammatical relations and ranked them by frequency (Kallas et al., 2015).

Currently, the database is being examined, edited and supplemented by lexicographers. One of the significant observations regarding editing collocations is that deleting is necessary mainly in the case of mistakes in tagging and due to insufficient disambiguation, but also in the case of specific terms that are not part of general purpose everyday Estonian. The analysis of extracted data revealed a significant number of terminological collocations that belong to different domains. The most frequent are the law, medical, mathematical, scientific, linguistic and sports domains.

Figure 3 illustrates how collocates are presented in the dictionary database. In the dictionary entry preview for the adjective *eitav* ‘negative’ there are three collocates that were automatically extracted and later (during the editing process) were manually identified as domain-specific collocations. These collocations are *eitav kõne* ‘negative’, *eitav kõneliik* ‘negative’ and *eitav lause* ‘negative sentence’. The domain label is KEEL ‘linguistics’.

kollokaat (collo)	<input type="text" value="kõne"/>	<p>2 eitav omadussõna {A} 3841</p> <p>Nimisõnaga modifies 2910 (7.547293) eitav vastus 1266 (7.134377) eitav seisukoht 425 (6.340386) eitav sulatumine 233 (6.538565) eitav otsus 203 (4.181146) eitav hoiak 109 (6.615781) eitav hinnang 100 (4.321532) eitav arvamus 38 (2.868828) eitav tulemus 35 (1.607514) eitav kõne KEEL 22 (2.949626) eitav positsioon 17 (3.249367) eitav lause KEEL 12 (2.665471) eitav kõneliik KEEL 8 (6.466184) eitav reaktsioon 8 (3.092124)</p>
järgnev märksõna		
vald	KEEL	
stiil		
kollok seletus		
kollok sagedus (freq)	22	
kollok skoor (score)	2.949626	
näidete grupp		
kol-näide (example)	Kui öeldis on &ba;eitavas&bl; kõnes, ei saa üldjuhul olla nimetavas ja omastavas käandes sihitist.	
kol-näide (example)	Aga töönaõse sündmuse mittetoimumist on otstarbekam teatada &ba;eitavas&bl; kõnes.	
kol-näide (example)	3. Isikulise tegumoe kindla kõneviisi &ba;eitavas&bl; kõnes koos	

Figure 3: An example of an entry for the adjective *eitav* ‘negative’ in DWS EELex: the editing window in XML view (left) and the dictionary entry preview (right)

In order to identify such collocations, different approaches are used: 1) consulting terminological dictionaries and databases, 2) analyzing available domain corpora, and 3) building new domain corpora within Sketch Engine with WebBootCaT (Baroni et al., 2006) and implementing the Term Extraction function (Kilgarriff et al., 2014; Fišer et al., 2016). The latter takes a lot of effort on the part of the lexicographer.

The automation of this task would have a major impact on lexicographic word sketch data analysis and (semi-)automated collocation dictionary compilation.

3. Multiword Term Extraction within Sketch Engine: State of the Art

In this section, we present the results of our experimental study on the reliability of the data that can be identified and extracted using methods that were developed within the Sketch Engine corpus query system, particularly the tools WebBootCaT (Baroni et al., 2006) and Term Extraction (Kilgarriff et al., 2014; Fišer et al., 2016). Term Extraction is based on comparing frequencies of pre-defined units in a domain corpus and a general corpus. The resulting term candidates are sorted by the ratio of the frequencies (the keyword score).

For the experiment, Russian and Estonian were used. Russian is highly represented on the Web (estimated percentage is 6.5%) while Estonian is not (estimated percentage is 0.1%).²

3.1 Term Grammar and Domain Corpora

Sketch Engine implements a data-driven approach to this problem: instead of having domain experts build such a lexicon from scratch using an automatic procedure that produces a high quality lexicon from the supplied domain-specific corpus. The whole procedure is described in detail in (Kilgarriff et al., 2014). Term candidates for a language domain can be found through the following steps:

- taking a corpus for the domain, and a reference corpus for the language;
- identifying the grammatical shape of a term in the language and writing a term grammar³;
- tokenizing, lemmatizing and POS-tagging both corpora;
- identifying (and counting) the items in each corpus which match the grammatical pattern;

² Accessed at: https://w3techs.com/technologies/overview/content_language/all (20 May 2017)

³ Term Grammar: Writing term grammar. Accessed at: <https://www.sketchengine.co.uk/documentation/writing-term-grammar/> (25 May 2017)

- for each item in the domain corpus, comparing its frequency with its frequency in the reference corpus.

The term identification is based on CQL—Corpus Query Language—to specify the term grammar for each language. The term grammar formalism can be defined as regular expressions over words, lemmas and morphological tags (imposing a requirement that the corpora be tagged). The format of the term grammar corresponds to the word sketch grammar and hence makes it possible to use the same indexing machinery for efficient storage and retrieval of the term candidates.

Altogether there are term definitions for 13 languages in Sketch Engine, Russian and Estonian among them. However, to the best of our knowledge, there are not many works dealing with the evaluation of these term grammars. The results of the evaluation presented in Fišer et al. (2016) were applied to the Slovene language. Adjective + noun combinations achieve 73% accuracy, whereas trigrams with prepositions have 63% accuracy.

The term grammars for Russian and Estonian were built on the assumption that terms are mostly noun phrases. This assumption is based on academic descriptions of term structures in Russian (Gerd, 1986) and Estonian (Erelt, 2007), and partly on the empirical observation of the terms structure in terminological databases (e.g., in the NATO English–Russian terminology lexicon⁴, out of 300 randomly chosen terms only two were verb phrases).

The Russian term definition consists of the following lexico-grammatical patterns (Khokhlova, 2009): 1) adjective + noun, 2) adjective + adjective + noun, 3) noun + noun, 4) noun + adjective, and 5) adjective + noun + noun. For Estonian, the patterns are: 1) adjective + noun, 2) noun + noun, and 3) noun + verb. Each model involves several restrictions on the grammatical forms of words.

For Russian, the terms are built on lemmas instead of word forms so that all of the flective variants contribute to the one lemmatized item.

For Estonian, colloc-type rules were used in order to extract multiword term candidates so that one component was presented as a lemma and the other one in the particular inflectional form, e.g. *sõjaväe konvoi* (the military-SG-GEN convoy-SG-NOM) ‘military convoy’.

In our experiment, as reference corpora we used large web corpora gathered using SpiderLing (Suchomel & Pomikalek, 2012). For Russian, this was Russian Web 2011 (ruTenTen11) and for Estonian Web 2013 (etTenTen13).⁵

⁴ NATO database: <http://www.nato.int/docu/glossary/eng/15-main.pdf> (20 May 2017)

⁵ Both corpora are available at <https://the.sketchengine.co.uk/auth/corpora/> (20 May 2017)

Domain corpora were built by WebBootCaT (Baroni et al., 2006), a tool for gathering domain specific documents from the web. As a domain corpus, we built a military corpus due to the good quality of military lexicons that can be used both for compiling such corpora and for evaluating term extraction. For Russian we used the NATO English–Russian terminology lexicon and for Estonian the database MILITERM⁶.

We used 145 monolexemic and multiword terms from the NATO list as seed words for the Russian military domain corpus. For example, *баллистическая ракета* ‘ballistic missile’, and *автоматическая система управления войсками* ‘automated command and control system’. The resulting size of the corpus was 25 million words.

We used 1500 monolexemic and multiword terms from MILITERM as seed words to build the Estonian domain corpus. For example, *õhusõidukite liikumise miinimumala* ‘minimum aircraft operating surface’ and *radarihävitaja* ‘wild weasel’. The resulting size of the corpus was only three million words. The reason for using a much higher count of seed terms compared to Russian was to get as many relevant texts from the web as possible. However, the resulting corpus was not big enough, as is shown in the evaluation.

To select the most relevant terms out of the term candidates set (with regard to the target domain), we compared their frequencies using the SimpleMaths method⁷ and computed a score for each term.

3.2 Evaluation and Discussion

We compared the extracted terms with the original terminology database and evaluated the recall of the whole WebBootCaT and Terminology extraction method.

The full terminological database was used for the evaluation. Since the seed words were a part of the full set they naturally occurred in the result domain corpus. The benefit of creating the domain corpus is that it also contains terms which were not used as seed phrases.

The evaluation showed that the task was a precision/recall tradeoff, as can be seen in Figures 4 and 5. Taking more candidates into account, the precision dropped while the recall grew. There were enough Russian web documents in the target domain found and downloaded to cover 50% of the single word terms and 25% of the multiword terms in the top 3,000 term candidates. Thanks to the size and the satisfactory representation of the target domain, the corpus can be used by

⁶ MILITERM database: <http://termin.eki.ee/militerm/> (20 May 2017)

⁷ <https://www.sketchengine.co.uk/documentation/simple-maths/> (20 May 2017)

lexicographers to study collocations of words from the domain. The same does not hold true for the Estonian corpus: it is too small and the target domain is poorly covered.

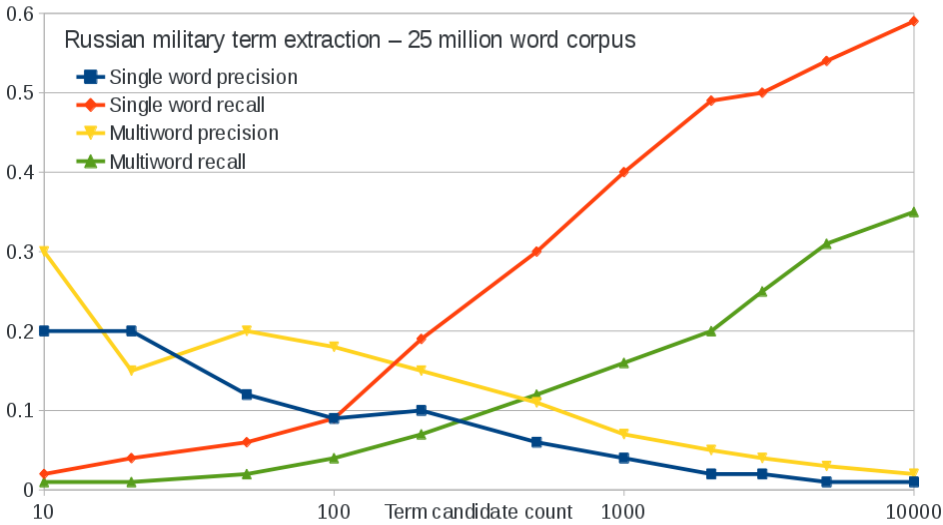


Figure 4: Evaluation of the top term candidates (with the highest keyword score) extracted from the Russian military domain corpus

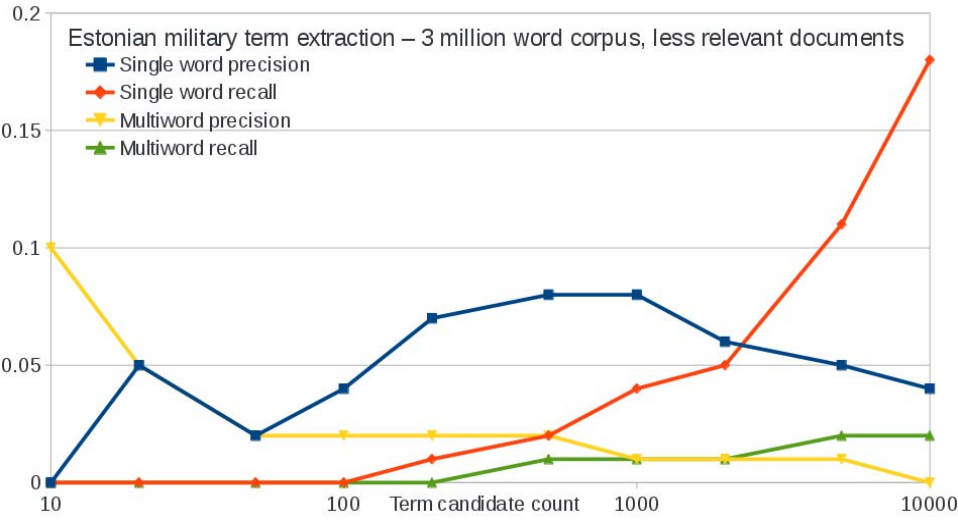


Figure 5: Evaluation of the top term candidates extracted from the Estonian military domain corpus

The most common reasons leading to a wrong classification in both languages were as follows:

- a term pattern not covered by the term grammar (e.g., more than five word terms or terms not consisting of noun phrases);
- a general noun phrase but not a term;

- a word or a phrase in the domain but not a good term;
- a part of a multiword term;
- valid terms from a different domain (e.g., politics rather than military in Estonian).

The experiment showed that this method works well only for languages that are highly represented on the Web and is insufficient for languages whose estimated percentages of the top 10 million websites is 0.1%. The result depends greatly on the size and quality of the domain corpus. The problem is that for languages with a small presence on the Web, the search engine cannot find enough documents in the domain. The minimum size for the domain corpus should be five or 10 million words.

4. Identification of Domain Preferences of Collocations in Word Sketches

In this section, we propose two possibilities for identification of domain preferences of collocations: 1) comparing frequency in a reference and a domain corpus to identify domain preferences of a headword and its collocates, and 2) comparing word sketches of reference and domain corpora (as an example see Figure 6).

The first approach requires domain corpora to compare frequencies of collocations in a domain and the focus corpus and display domain preferences of headwords and collocations in a way similar to the indication of register preference in Figure 2. In general, any document attribute that is relevant for lexicography could be used to define a subcorpus of the focus corpus. If a collocation was mainly found in a single subcorpus based on the selected document attributes, it would be labelled by the corresponding text type in the word sketch interface. For example, taking advantage of language variety, genre and topic subcorpora, word ‘lamer’⁸ could be labelled ‘Usually American English, Internet forum, Computers’ which constitutes valuable information for a lexicographer.

The second approach suggests that another possible way to analyze the domain preference of collocations is to implement the procedure used in Bilingual Word Sketch function⁹ (Kovář, Baisa & Jakubíček, 2016). Figure 6 illustrates the sketch for the word *операція* ‘operation’, where adjectival collocates from a reference corpus and from a domain corpus are presented.

⁸ <https://en.oxforddictionaries.com/definition/us/lamer> (10 July 2017)

⁹ <https://www.sketchengine.co.uk/user-guide/user-manual/bilingual-word-sketch/> (20 May 2017)

операция (noun)
 Russian Web 2011 freq = **217,688** (173.60 per million)

операция NATO Terms Russian freq = **20,207** (621.55 per million)

a modifier	28,079	12.90	a modifier	5,985	29.62
пластический	<u>1,152</u>	9.94	наступательный	<u>491</u>	11.17
пластической операции			наступательная операция ,		
контртеррористической	<u>640</u>	9.50	десантный	<u>248</u>	10.16
контртеррористической операции			десантной операции		
хирургический	<u>887</u>	9.23	контртеррористической	<u>140</u>	9.54
хирургической операции			контртеррористической операции		
спасательный	<u>428</u>	8.68	оборонительный	<u>166</u>	9.43
спасательной операции			оборонительная операция ,		
наступательный	<u>297</u>	8.29	антитеррористической	<u>121</u>	9.32
наступательная операция			антитеррористической операции в		
антитеррористической	<u>236</u>	8.00	специальный	<u>177</u>	8.85
антитеррористической операции в			специальной операции		
миротворческий	<u>217</u>	7.88	берлинский	<u>82</u>	8.73
миротворческой операции			Берлинской операции		
десантный	<u>209</u>	7.81	гуманитарный	<u>90</u>	8.67
десантной операции			гуманитарной операции		
профилактический	<u>298</u>	7.78	воздушно-десантная	<u>77</u>	8.64
профилактической операции «			Вяземская воздушно-десантная операция		
полостной	<u>185</u>	7.73	предстоящий	<u>77</u>	8.55
полостной операции			предстоящей операции .		
срочный	<u>312</u>	7.70	карательный	<u>69</u>	8.50
срочная операция			карательной операции		
военный	<u>1,388</u>	7.63	миротворческий	<u>67</u>	8.41
военной операции			миротворческой операции в		

Figure 6: Word sketch for the noun операция ‘operation’ with aligned grammatical relations in the Russian Web 2011 corpus and the NATO Terms Russian domain corpus

The first three collocates in the reference corpora are *пластический* ‘plastic surgery’, *контртеррористический* ‘counterterrorist (operation)’, and *хирургический* ‘surgical (operation)’. The most frequent collocates in the domain corpora are *наступательный* ‘offensive (operation)’, *десантный* ‘amphibious (operation)’, and *контртеррористический* ‘counterterrorist (operation)’. This helps to separate collocations and the word sense associated to a single topic represented by the military domain corpus.

5. Conclusion and Future Work

The results of our experiment revealed that for languages that are highly represented on the Web it is possible to create sizable domain corpora. We propose to exploit the domain corpora for automatic comparison of frequencies of collocations in a domain and a reference corpus to help lexicographers by indicating domain preferences of words and their collocates.

Our study can be implemented to improve the efficiency of word sketch data analysis and it is important to stress that the procedure itself is not language-specific, but depends on how highly a language is represented on the Web. The components required include a reference corpus, a number of different domain corpora (a minimum of 5 to 10 million words), a Sketch Grammar and a Term Grammar.

We suggest possible methodological improvements for corpus tools in order to improve automatic and semi-automatic collocations dictionary compilation by automatic indication of domain preferences. Domain preference provides useful information to users and allows to distinguish terminological collocations.

6. References

- Atkins, S.B.T. & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press.
- Costa, R. & Silva, R. (2004). The Verb in the Terminological Collocations Contribution to the Development of a Morphological Analyser MorphoComp. *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26-28, 2004, Lisbon, Portugal*. European Language Resources Association.
- Erelt, T. (2007) *Terminiõpetus*. Tartu: Tartu Ülikooli kirjastus.
- Fišer, D., Suchomel, V., & Jakubíček, M. (2016). Terminology Extraction for Academic Slovene Using Sketch Engine. *Tenth Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2016*. Brno: Tribun EU, pp. 135–141.
- Frantzi, K., Ananiadou, S., & Mima, H. (2000). Automatic recognition of multi-word terms: the C-value/NC-value method. *International Journal on Digital Libraries*, 3(2), pp. 115–130.
- Gerd, A. (1986) *Osnovy naučno-tekničeskoj leksikografii*. Leningrad: izd-vo LGU.
- IATE: The EU's multilingual term base. Accessed at: <http://iate.europa.eu>. (25 May 2017)
- Kallas, J., Kilgarriff, A., Koppel, K., Kudritski, E., Langemets, M., Michelfeit, J., Tuulik, M., & Viks, Ü. (2015). Automatic generation of the Estonian Collocations Dictionary database. *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*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd, pp. 1–20.
- Kilgarriff, A., Jakubíček, M., Kovář, V., Rychlý, P. & Suchomel, V. (2014). Finding Terms in Corpora for Many Languages with the Sketch Engine. *Proceedings of the Demonstrations at the 14th Conference the European Chapter of the Association for Computational Linguistics*. Sweden, pp. 53–56
- Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The sketch engine. *Proceedings EURALEX 2004*, Lorient, France, pp. 105–116.
- Khokhlova, M. (2009). Applying Word Sketches to Russian. *Proceedings of Raslan 2009. Recent Advances in Slavonic Natural Language Processing*. Brno: Masaryk University, pp. 91–99.
- Kovář, V., Baisa, V. & Jakubíček, M. (2016). Sketch Engine for Bilingual lexicography. *International Journal of Lexicography*, 29(3), pp. 339–352.

- OCDSE: Oxford collocations dictionary for students of English. (2002). Oxford: Oxford University Press.
- Ramisch, C. (2009). *Multi-word terminology extraction for domain-specific documents*. Master's thesis, École Nationale Supérieure d'Informatique et de Mathématiques Appliquées, Grenoble, France. Accessed at: http://www.inf.ufrgs.br/~ceramisch/download_files/publications/2009/p01.pdf (25 May 2017)
- Rundell, M. (2012). The road to automated lexicography: an editor's viewpoint. In S. Granger & M. Paquot (eds) *Electronic Lexicography*. Oxford: Oxford University Press, pp. 15-30.
- Rundell, M. & Kilgarriff, A. (2011). *Automating the creation of dictionaries: where will it all end?* In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (eds) *A Taste for Corpora. A tribute to Professor Sylviane Granger*. Benjamins. P., pp. 257-281.
- Vainik, E. (1999). Millest on tehtud õigusterminid? *Õiguskeel*, pp. 27–39.
- Sanjuan, E., Dowdall, J., Ibekwe-SanJuan, F., & Rinaldi, F. (2005) A symbolic approach to automatic multiword term structuring. *Computer Speech & Language Special Issue on Multiword Expressions*, 19(4), pp. 524–542.
- SkELL: Sketch Engine for Language Learning. Accessed at: <http://skell.sketchengine.co.uk/run.cgi/skell>. (25 May 2017)
- Svensen, B. (2009). *A handbook of lexicography. The theory and practice of dictionary-making*. Cambridge: Cambridge University Press.

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

