

# Meaning structure of cognate words in English and Russian: comparing word sense frequency\*

Boris Iomdin<sup>1,2</sup>, Konstantin Lopukhin<sup>3</sup>, Anastasiya Lopukhina<sup>1</sup>,  
Grigory Nosyrev<sup>4</sup>

<sup>1</sup>Vinogradov Russian Language Institute, Russian Academy of Sciences, Moscow, Russia

<sup>2</sup>School of linguistics, National Research University Higher School of Economics

<sup>3</sup>Scrapinghub, Moscow, Russia

<sup>4</sup>Yandex, Moscow, Russia

## Abstract

Polysemy is a key issue in theoretical semantics and lexicography as well as in computational linguistics. When words have several senses, it is important to describe them properly in the dictionary (a lexicographic task) and to be able to distinguish between them in given context (a computational linguistics task, known as WSD). Recently attention has been drawn to the fact that different senses normally have different frequencies in corpora. Elsewhere we reported on our research into that issue and introduced several techniques for determining sense frequency based on dictionary entries matched with data from large corpora. Information about word sense frequency may enrich language learning resources and help lexicographers order senses within a word according to frequency, if needed. When learning a foreign language, a student may encounter a word that exists in his/her native language (as a borrowing or an international word), and is tempted to assume that the foreign word and its equivalent have the same meaning structure. However, sometimes this is not the case, and the most frequent sense of a word in one language may be much less frequent for its cognate. We propose a method for detecting such cases. For that purpose, we selected a set of Russian words included into the Active Dictionary of Russian, which have more than two dictionary senses and have cognates in English. We estimated frequencies for English and Russian senses using SemCor and Russian National Corpus respectively, matched senses in each pair of words and compared their frequencies. In this way, we revealed cases in which the most frequent senses and the whole meaning structures are, cross-linguistically, substantially different and studied them in more detail. As a result, we obtained information that may prove useful for learners of Russian or English as well as for lexicographers and computational linguists dealing with machine translation.

**Key words:** lexicography, semantics, meaning, frequency, sense frequency, WSD, semantic vectors, text corpora, parallel corpora, cognates, English lexicon, Russian lexicon

## 1. Introduction

When words have several senses, it is important to describe them properly in the dictionary and to be able to distinguish between them in any given context (see e.g. Pustejovsky, 1996; Apresjan, 2000; Lin and Ahrens, 2005; Agirre and Edmonds, 2007; Kwong, 2012; Hanks, 2013; Iomdin, 2014). Although polysemy is a key issue in theoretical semantics, lexicography and computational linguistics, sense frequency distributions are rarely the focus of researchers' attention, as it is quite difficult to obtain such information. Recently, attention has been drawn to the fact that different senses of a word normally have different frequencies in corpora (see e.g. Iomdin, Lopukhina and Nosyrev, 2014). However, there are almost no resources that provide this data.

---

\*The research of Boris Iomdin, Konstantin Lopukhin and Anastasiya Lopukhina was supported by RSF (project No.16-18-02054: Semantic, statistic and psycholinguistic analysis of lexical polysemy as a component of Russian linguistic worldview). Parsing the Dante database of English was done by Grigoriy Nosyrev.

The question of the most frequent (or predominant) sense (MFS) has been discussed for the purposes of automated word sense disambiguation task (WSD), as MFS is considered to be a very powerful heuristic, which is difficult to overcome for many WSD systems (Ide and Véronis, 1998; Navigli, 2009). For English, several approaches to acquiring predominant senses have been applied so far. Mohammad and Hirst's (2006) method makes use of the category information in the Macquarie Thesaurus; McCarthy et al. (2007) proposed an unsupervised approach for finding the predominant senses using a distributional thesaurus; Bhingardive et al. (2015) compared the embedding of a word with all its sense embeddings (which are produced using various features of WordNet) and obtained the predominant sense with the highest similarity. For Russian, a pilot study of MFS detection was presented by Loukachevitch and Chetviorkin (2015), who used the Thesaurus of Russian Language (RuThes-lite) to determine the most frequent sense of ambiguous nouns, verbs and adjectives with the help of monosemous multiword expressions that are related to those words. Their results are comparable to the state-of-the-art in this field – the highest accuracy rate reaches 57.4%.

The overall sense distribution is a question that is rarely put in focus. For English nouns, Lau et al. (2014) proposed a topic modeling-based method of estimating word sense distributions, based on Hierarchical Dirichlet Processes and on word sense induction, probabilistically mapping automatically learned topics to senses in a sense inventory. Some information about English verb pattern frequency distributions can be found in the Pattern Dictionary of English Verbs, developed by Patrick Hanks and colleagues (<http://pdev.org.uk/>; Hanks and Pustejovsky, 2005; Hanks, 2008). The authors emphasize that senses are associated with patterns (collocations) and not with words and that the Pattern Dictionary provides information about the relative frequency of phraseological patterns rather than dictionary senses. Cf. also Gries et al. 2010, where frequency distributions of English verbal constructions are discussed.

For Russian, a word sense frequency acquisition method and its evaluation for nouns were presented in (Lopukhina, Lopukhin and Nosyrev, submitted) and (Lopukhina et al., 2016). In these contributions we reported on our research into this issue and introduced several techniques for determining sense frequency based on dictionary entries matched with data from large corpora. This method is based on building context representations with semantic vectors and gives robust frequency estimates with little annotated examples available from dictionaries. Supplied with examples and definitions from the Active Dictionary of Russian (Apresjan et al., 2014), the method achieves frequency estimation error of 11-15% without any additional labeled data. It was used to obtain sense frequencies of 440 polysemous Russian nouns. Sense frequencies are estimated by performing WSD on a large number of contexts randomly sampled from corpora. WSD technique is refined to account for the case of few examples available. Context vectors are calculated by taking a weighted average of individual word vectors, and the sense vector is just an average of context vectors of all dictionary examples for this sense. When performing disambiguation, each context from corpora is assigned to a sense vector closest in context vector space. This method is able to achieve 75-77% disambiguation accuracy with a small number of examples available from the Active Dictionary of Russian (details in Lopukhina, Lopukhin and Nosyrev, submitted; Lopukhina et al., 2016).

## 2. Meaning frequency and foreign language learning

Information about word sense frequency may not only help lexicographers reasonably order senses within a lexical entry, taking account of their frequencies as one of the leading factors, but also enrich language learning resources. When learning a foreign language, a student may encounter a word that exists in his/her native language (as a borrowing or an international word), and is tempted to assume that the foreign word and its equivalent have the same meaning structure. However, often this is not the case.

In 1928, Koessler and Derocquigny coined the term “faux amis” for pairs of identical or similar cognate words with different senses, emphasizing the importance of such pairs to be identified and properly described so that interpreters and translators could avoid using such words erroneously (Koessler and Derocquigny, 1961). The authors exemplified the false friends by English and French *humour*, *application*, *distraction*, *gallant*, *attend* vs. *attendre*, *deception* vs. *déception*, etc. Ample lexicographic literature is devoted to such cases in different languages (see e.g. Darbelnet, 1970; Walter, 2002; Szpila, 2006; Vrbinc and Vrbinc, 2014, to name but a few). In many cases, cognates in a language pair do have some senses in common, but the meaning structure is different. An important case for language learners is when the most frequent sense of a word in one language turns out to be much less frequent for its cognate. Hence, a technique that could help identifying such words would undoubtedly be useful.

## 3. Material

Our primary word sense inventory and source of materials for Russian was the Active Dictionary of Russian (ADR), an ongoing project led by Juri Apresjan (see Apresjan et al., 2014). ADR is the first attempt at creating a full-fledged production dictionary of the Russian language. The main unit of the ADR, the lexeme, is a well-established word sense identified by a set of its unique properties (syntactic, semantic and pragmatic features, sets of synonyms, analogues, antonyms and semantic derivatives etc.). The lexical entries for all lexemes contain many various usage examples based on large corpora (mainly the Russian National Corpus), which is crucial for studying of sense frequencies.

We selected a set of Russian nouns included into the first two volumes of ADR, which have more than two dictionary senses and have cognates in English. Below is the list of these nouns:

автомат	automaton	арка	arc	батарея	battery
автономия	autonomy	армия	army	бизнес	business
автор	author	аромат	aroma	блок	block
авторитет	authority	артист	artist	блокада	blockade
агент	agent	архитектор	architect	брат	brother

MEANING STRUCTURE OF COGNATE WORDS IN ENGLISH AND RUSSIAN: COMPARING FREQUENCY

агентство	agency	архитектура	architecture	бригада	brigade
агрессия	aggression	ассистент	assistant	бутылка	bottle
адвокат	advocate	ассоциация	association	буфет	buffet
администрация	administration	асфальт	asphalt	бюллетень	bulletin
адрес	address	атака	attack	вагон	wagon
акт	act	атмосфера	atmosphere	гавань	haven
акцент	accent	аудитория	auditorium	галерея	gallery
акция	action	багаж	baggage	гараж	garage
алкоголь	alcohol	база	base	гарантия	guarantee
алтарь	altar	баланс	balance	гардероб	wardrobe
альбом	album	балет	ballet	гармония	harmony
анализ	analysis	балкон	balcony	гель	gel
ангел	angel	банда	band	гид	guide
анекдот	anecdote	бандит	bandit	гимн	hymn
аппарат	apparatus	банк	bank	горизонт	horizon
аргумент	argument	бар	bar	градус	grade
арена	arena	барьер	barrier	группа	group
арест	arrest	бас	bass	гусь	goose

Some of these words are true cognates of the same Indo-European origin (*брат* — *brother*, *гусь* — *goose*), some are English borrowings (*бар* < *bar*, *бизнес* < *business*), while many of them were borrowed from French or German and mostly have the same Latin or Greek roots. Etymology, however, is not relevant for this study, so we will refer to all word pairs under discussion as cognates.

Sense frequencies of the Russian nouns were estimated automatically by performing word sense disambiguation on contexts sampled from corpus and then calculating relative sense frequencies in the sample. For the purposes of the current study, we have sampled 1000 random contexts for each word from a domain-neutral Russian National Corpus (RNC, ruscorpora.ru, 230 million tokens in the main corpus), a resource made by a consortium of linguists and developers and considered to be the best academic corpus for Russian. In (Lopukhina, Lopukhin and Nosyrev, submitted) we compared sense frequency distributions in RNC and web-based RuTenTen11 (Kilgarriff et al., 2004) and found out that despite their

differences both corpora have similar sense frequency distributions and overall statistics (e. g. 82% of nouns have the same most frequent sense).

Sense frequencies of the English counterparts were obtained from the largest sense tagged SemCor 3.0 corpus (Miller et al., 1993). SemCor is composed of 220,000 words taken from the Brown corpus (Francis and Kučera, 1979). Approximately half of the words in this corpus are open-class words (nouns, verbs, adjectives and adverbs), which have been linked to WordNet 3.0 senses (Fellbaum, 1998) by human taggers using a software interface. SemCor (and WordNet-like resources, in general) is often criticized for its excessively fine-grained sense distinction that is not supported by syntactic, syntagmatic or semantic criteria, is not really needed for NLP tasks (Hanks and Pustejovsky, 2005; Navigli, 2006; Snow et al., 2007) and does not reflect the way people represent word meaning (Ide and Wilks, 2007; Brown, 2008). Nonetheless, SemCor remains the state-of-the-art resource in most WSD experiments. For the purposes of this study, we selected a subset of words that occur at least 20 times in SemCor 3.0 and calculated the frequencies of their senses directly from labeled SemCor contexts, with an estimated maximum frequency error of 15-20%.

The following table shows a sample entry of the proposed bilingual Russian-English cognates dictionary with senses frequencies (for the word pair *авторитет* — *authority*). The senses are ordered according to their frequency in RNC and paired with their English counterparts, when available; then the senses of the English word that have no matches in Russian are listed according to their frequency in SemCor. For senses that only exist in one of the two languages, a translation equivalent is given in parentheses. The most frequent senses in both languages are printed in boldface. The explications of the senses in Russian are taken from ADR, the explications of the senses in English are adapted from either WordNet or the MacMillan Online Dictionary (<http://www.macmillandictionary.com>).

<i>авторитет</i>		<i>authority</i>	
<p><b>‘Свойство человека, группы людей или организации A1, состоящее в том, что люди A3 считают нужным учитывать мнение A1 об A2, потому что A1 хорошо знает область деятельности A2’</b></p> <p>[‘The property of a human, a group of people or an organization A1, such that people A3 see it proper to take into account the opinion of A1 on A2 because A1 knows well the field A2’].</p>	<b>0.68</b>	<p>‘Freedom from doubt; belief in yourself and your abilities’.</p>	0.067
‘Человек, имеющий авторитет в	0.28	‘An expert whose views are	0.1

сфере А2'. [‘Someone who has such property in field А2’]		taken as definitive.’	
‘Человек, имеющий авторитет в уголовном мире’. [‘Someone who has such property in criminal underworld’]	0.04	( <i>crime boss</i> )	—
( <i>власть</i> )	—	‘The power or right to give orders or make decisions’.	<b>0.6</b>
( <i>власти</i> )	—	‘Persons who exercise (administrative) control over others’.	0.167
( <i>управление</i> )	—	‘An administrative unit of government’.	0.033
( <i>полномочия</i> )	—	‘Official permission or approval’.	0.033

#### 4. Discussion

Comparing the senses of cognates by taking into account their frequency, one can detect various cases of cognates whose meaning structures are dissimilar. For some words, no senses have a match in the other language at all (these are authentic “faux amis”). In our data, typical examples of such pairs are *арка* (most frequent sense ‘a structure with a curved top and two straight sides that you can walk through, an arch’) vs. *arc* ‘a curved shape’ (Fig. 1),



Figure 1. *Арка* and *arc*

*вагон* ‘railway carriage’ vs. *wagon* ‘any of various kinds of wheeled vehicles drawn by an animal or a tractor’, *градус* ‘degree’ vs. *grade* ‘a level of school’. In other cases, there is one matching sense (or more), but the most frequent senses differ drastically. Apart from the example of *авторитет* vs. *authority* given above, cf. *акция* ‘one of the equal parts of a company that you can buy, a share’ vs. *action* ‘something done’, *армуц* ‘someone who

performs in plays and films’ vs. *artist* ‘someone who makes paintings, sculptures etc’, *банда* ‘a group of criminals acting together’ vs. *band* ‘a group of musicians’.

Even harder for learners are pairs where several senses match but others do not (and learners naturally tend to use them erroneously). Cf. *бар* vs. *bar* (senses ‘a rigid piece of metal or wood’, ‘an obstruction placed at the top of a goal’, ‘the act of preventing’ absent in Russian, sense ‘liquor cabinet’ absent or very rare in English); *база* vs. *base* (senses ‘lowest support of a structure’, ‘the part of an organ nearest its point of attachment’ absent in Russian, sense ‘camping site’ absent in English); *блок* vs. *block* (sense ‘a rectangular area in a city surrounded by streets’ absent in Russian).

Some of the mismatches exemplified above are not mentioned in dictionaries, and lack of knowledge thereof often leads to erroneous translations or usage. E.g. the expression *criminal authority* can be found in genuine English texts, meaning ‘power to make decisions with regard to crime’ (cf. *State officials acknowledged they did not recover all revoked permits and the state had no civil or criminal authority to force the surrender of revoked permits*), but it also quite frequently occurs in texts translated from Russian and meaning ‘crime boss’, cf. *The Russian mass media informed that Alexander Matusov nicknamed Basmach, head of the Shelkov criminal gang, a criminal authority, has been arrested in Thailand*. The expression *touristic base* normally means ‘a group of touristic businesses that form an important part of an economy’ (cf. *Glasgow has greatly expanded the touristic base of the region*), but it can also be found in texts apparently written by non-natives, where it means ‘a camping site’ (cf. *The touristic base built in Erzhei for the 'Oktai' ensemble is now open not only to the young singers but to everybody*).

Online dictionaries and translation memories dealing with parallel corpora sometimes contain such texts, too, which can result in misleading their users. E.g. the recently opened resource *linguee.com*, a powerful translation tool combining an editorial dictionary and a search engine for parallel corpora, provides 28 examples of parallel Russian-English texts for the Russian word *вагон*, in 5 of which it is rendered as *wagon* in English (all of them taken from Russian or Czech websites and clearly representing translations into English rather than genuine English texts), cf. *Once, when they were travelling by train, a wagon accidentally disconnected from the train and began to roll slowly down a slope* (from *Skolkovo.ru*). An inverted example: in the same dictionary we can see the English word *arc* translated into Russian as *арка* in the following sentence: *The result will be an arc defined by three points — В результате получится арка, построенная по трем точкам* (in this geometrical context, one should use *дуга* rather than *арка*).

Also, usage of cognate words and their representation in the mental lexicon naturally differs in monolingual and bilingual speakers and communities (see e.g. Schreuder and Weltens, 1993; Jiang, 2004; Dong et al., 2005; Degani and Tokowicz, 2013). By distinguishing senses that are not shared in cognate pairs in standard language, we can more easily reveal cases when they are mixed up and include them as examples of non-standard usage into standard language learning manuals or dictionaries. For example, the word *блок* is widely used by Russian immigrants in the USA in the sense of ‘the distance along a city street from where one road crosses it to the next road’ (absent in standard Russian, where its equivalent is the

word *квартал*): *С авеню надо свернуть на улицу и проехать два блока* ‘One has to turn to the street from the avenue and to drive two blocks’, *Я знаю отличный бар в паре блоков отсюда* ‘I know a swell bar a couple of blocks from here’.

Hence it seems reasonable to start creating resources where for all cognates and pairs of seemingly identical words in different languages, all their senses would be compared and supplied with frequency information in monolingual corpora.

## 5. Conclusion

In this study, we performed a pilot experiment, determining sense frequencies for cognate Russian and English words in corpora and comparing their meaning structures taking this information into account. The results seem quite promising: we can obtain data that may prove useful for learners of Russian or English as well as for lexicographers and computational linguists dealing with machine translation or deep semantic analysis.

The main issue now is the lack of large semantically annotated corpora and dictionaries with a sufficient number of examples for each word sense, which limits the possibilities of automatic techniques for calculating meaning frequency. Our future plans include the following initiatives:

- apply the method of estimating word sense frequencies used for Russian on the base of the examples provided in the Active Dictionary of Russian (Lopukhina, Lopukhin, Nosyrev, submitted; Lopukhina et al., 2016) to English, using data of the MacMillan dictionary and the Dante database ([www.webdante.com](http://www.webdante.com));
- create a database of Russian-English cognates comparing their senses according to their frequency;
- study parallel Russian-English corpora (primarily subcorpora of Russian National Corpus) to investigate into possible differences in meaning structures of cognates used by native and non-native speakers, in Russian-to-English and English-to-Russian translations;
- expand the method to other language pairs; inter alia, compare closely related languages (such as Russian and Polish) to provide material for comparative semantic and lexicological studies.

## References

- Agirre E. and Edmonds P. (Eds.). (2007). *Word sense disambiguation: Algorithms and applications* (Vol. 33). Springer Science & Business Media.
- Apresjan J. (ed.) (2014). *Active Dictionary of Russian*. Vol. 1: A-B, Vol 2: V-G. Moscow: Jazyki slavjanskix kul'tur.
- Apresjan J. (2000). *Systematic Lexicography*. Oxford.
- Bhingardive S., Dhirendra S., Murthy V. R., Hanumant R. and Pushpak B. (2015). Unsupervised Most Frequent Sense Detection using Word Embeddings. *North American Chapter of the Association for Computational Linguistics (NAACL)*, Denver, Colorado.



- Brown S. W. (2008). Choosing sense distinctions for WSD: Psycholinguistic evidence. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*. Association for Computational Linguistics, pages pp. 249-252.
- Darbelnet J. (1970). Dictionnaires bilingues et lexicologie différentielle. *Langages*, 19:92-102.
- Degani T. and Tokowicz N. (2013). Cross-language influences: Translation status affects intraword sense relatedness. *Memory & cognition*, 41(7):1046-1064.
- Dong Y., Gui S. and MacWhinney B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition*, 8(03):221-238.
- Fellbaum C. (ed.) (1998). *WordNet, An Electronic Lexical Database*. The MIT Press, Cambridge, MA.
- Francis W. N. and Kučera H. (1979). *A Standard Corpus of Present-Day Edited American English, for use with Digital Computers (Brown)*. Brown University. Providence, Rhode Island.
- Gries S. T., Hampe B. and Schönefeld D. (2010). Converging evidence II: More on the association of verbs and constructions. In: *Empirical and experimental methods in cognitive/functional research*, CSLI Publications, pages 59-72.
- Hanks P. (2008). Mapping meaning onto use: a Pattern Dictionary of English Verbs. In *Proceedings of the ACL*, Utah.
- Hanks P. (2013). *Lexical analysis: Norms and exploitations*. Boston: MIT Press.
- Hanks P. and Pustejovsky J. (2005). A Pattern Dictionary for Natural Language Processing. *Revue Française de linguistique appliquée*, 10(2):63-82.
- Ide N. and Véronis J. (1998). Introduction to the special issue on word sense disambiguation: the state of the art. *Computational linguistics*, 24(1):2-40.
- Ide N. and Wilks Y. (2007). Making sense about sense. In *Word sense disambiguation*. Springer Netherlands: 47-73.
- Iomdin B. (2014). Polysemous words in and out of the context. *Voprosy Jazykoznanija*, 4:87-103.
- Iomdin B., Lopukhina A., Nosyrev G. (2014). Towards a word sense frequency dictionary. In: *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2014"*, pages 205-219.
- Jiang N. (2004). Semantic transfer and development in adult L2 vocabulary acquisition. In *Vocabulary in a second language: Selection, acquisition, and testing*, pages 101-126.
- Kilgarriff A., Rychly P., Smrz P. and Tugwell D. (2004). The Sketch Engine. In *Information Technology Research Institute Technical Report Series*, pages 105-116.
- Koessler M. and Derocquigny J. (1961). *Les faux amis ou les pièges du vocabulaire anglais*. Paris: Vuibert.
- Kwong O. Y. (2012). *New perspectives on computational and cognitive strategies for word sense disambiguation*. New York: Springer Science & Business Media.
- Lau J. H., Cook P., McCarthy D., Gella S. and Baldwin T. (2014). Learning word sense distributions, detecting unattested senses and identifying novel senses using topic models. In *Proceedings of ACL*, pages 259-270.
- Lin C. C. and Ahrens K. (2005) How many meanings does a word have? Meaning estimation in Chinese and English. In *Language acquisition, change and emergence: Essays in evolutionary linguistics*. Hong Kong, pages 437-464.
- Lopukhina A., Lopukhin K. and Nosyrev G. Automated word sense frequency estimation for Russian nouns. In *Quantitative Approaches to the Russian Language*. In print. Available online: [sensefreq.ruslang.ru/download/Automated\\_Word\\_Sense\\_Frequency\\_Estimation\\_for\\_Russian\\_Nouns\\_Lopukhina\\_et\\_al.pdf](http://sensefreq.ruslang.ru/download/Automated_Word_Sense_Frequency_Estimation_for_Russian_Nouns_Lopukhina_et_al.pdf).

- Lopukhina A., Lopukhin K., Iomdin B. and Nosyrev G. (2016). The taming of the polysemy: automated word sense frequency estimation for lexicographic purposes. In *Proceedings of EURALEX-2016*. In print.
- Loukachevitch N. and Chetviorkin I. (2015). Determining the Most Frequent Senses Using Russian Linguistic Ontology RuThes. In *Proceedings of the workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities at NODALIDA*.
- McCarthy D., Koeling R., Weeds J. and Carroll J. (2007). Unsupervised acquisition of predominant word senses. *Computational Linguistics*, 33(4):553-590.
- Miller G. A., Leacock C., Tengli R. and Bunker R. T. (1993). A semantic concordance. In *Proceedings of the workshop on Human Language Technology, Association for Computational Linguistics*, pages 303-308.
- Mohammad S. and Hirst G. (2006). Determining Word Sense Dominance Using a Thesaurus. In EACL.
- Navigli R. (2006). Meaningful clustering of senses helps boost word sense disambiguation performance. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics*, pages 105-112.
- Navigli R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR)* 41.2: 10.
- Pustejovsky J. (1996). *Lexical semantics: The problem of polysemy*. Oxford.
- Schreuder R. and Weltens B. (Eds.). (1993). *The bilingual lexicon* (Vol. 6). John Benjamins Publishing.
- Snow R., Prakash S., Jurafsky D. and Ng A. Y. (2007). Learning to merge word senses. In *Proceedings of EMNLP*. Prague, Czech Republic.
- Szpila G. (2006). False friends in dictionaries. Bilingual false cognates lexicography in Poland. *International Journal of Lexicography*, 19(1):73-97.
- Vrbinc M. and Vrbinc A. (2014). Friends or Foes? Phraseological False Friends in English and Slovene. *AAA: Arbeiten aus Anglistik und Amerikanistik*, pages 71-87.
- Walter H. (2002). Les "faux amis" anglais et l'autre côté du miroir. *La linguistique*, 37(2):101-112.