Automatic Categorisation of Multiword Expressions and Named Entities in Bulgarian

Ivelina Stoyanova
DCL, IBL, BAS
Sofia, Bulgaria
iva@dcl.bas.bg

Abstract

This paper describes an approach for automatic categorisation of various types of multiword expressions (MWEs) with a focus on multiword named entities (MNEs), which compose a large portion of MWEs in general. The proposed algorithm is based on a refined classification of MWEs according to their idiomaticity.

While MWE categorisation can be considered as a separate and independent task, it complements the general task of MWE recognition. After outlining the method, we set up an experiment to demonstrate its performance. We use the corpus Wiki1000+ that comprises 6,311 annotated Wikipedia articles of 1,000 or more words each, amounting to 13.4 million words in total. The study also employs a large dictionary of 59,369 MWEs noun phrases (out of more than 85,000 MWEs), labelled with their respective types. The dictionary is compiled automatically and verified semi-automatically.

The research presented here is based on Bulgarian although most of the ideas, the methodology and the analysis are applicable to other Slavic and possibly other European languages.

1. Introduction

Statistical analyses show that multiword expressions (MWEs) comprise a significant part of the lexical system of a language. For instance, 24.49% of the Bulgarian WordNet and 22.5% of the Princeton WordNet 2.0 (Koeva, 2006) are MWEs. MWEs pose a complex set of problems to both theoretical linguistics and Natural Language Processing (NLP). Developing efficient methods for their automatic identification and categorisation will help improve results in Information Retrieval, Machine Translation, and other areas of Computational Linguistics.

A wide variety of approaches towards MWE recognition have been developed in recent years. Generally, they differ in the amount of linguistic information used and the particular statistical tools applied in the analysis. However, neither statistical methods nor methods heavily dependent on linguistic resources have proved successful for the general purpose of MWE recognition independently of each other, which has led to extensive exploration of hybrid methods.

Moreover, MWEs exhibit a wide variety of features and types, which additionally complicates their automatic processing. This paper presents an approach towards the automatic categorisation of MWEs following their automatic recognition. Multiword named entities (MNE) comprise a large portion of MWEs and are thus paid special attention here.

The research presented in this paper is based on Bulgarian although the methodology and analysis are largely applicable to other Slavic languages and possibly to other European languages as well.

2. Characteristics of MWEs

2.1. Main Features

The classification of MWEs we employ uses the feature *idiomaticity* in the sense of Nunberg et al. (1994), who consider this to be a chief characteristic of MWEs. It combines the degree of conventionality, understandability and compositionality of the MWE. Baldwin (2006) discusses a similar characteristic of MWEs and proposes a complex model for description of lexical units based on the following types of markedness:

- *Lexical markedness* – lexical and grammatical constraints on the realisation, such as paradigmatic constraints, e.g. *kick the bucket* but not *kick the buckets*, prosodic markedness, etc.;
- *Syntactic markedness* – syntactic irregularities in gender agreement or lack of agreement, or institutionalisation where lexemes preserve their historical characteristics regardless of the changes in the modern language, e.g. the preservation of the masculine gender of the noun *vecher* ('evening'), a feminine noun in modern Bulgarian, in the expression *Dobar vecher* ('Good evening').
- *Semantic markedness* – a relative (non-)compositionality of meaning, semantic relations (such as synonymy) with single words, e.g. *poshtenska stantsiya – postha*, both meaning 'a post office';
- *Pragmatic markedness* – in cases where the pragmatic features of the MWE components differ from those of the MWE as a whole, or the MWE is associated with a particular pragmatic reference point – consider the expression *Pusheneto zabraneno!*, literally *Smoking forbidden!* ('No smoking!') which is appropriate in certain communicative situations and not suitable in others;
- *Statistical markedness* – conventionality is reflected by high frequency of occurrence of particular collocations and markedly low or zero frequency of its synonymous counterparts, e.g. *strogo sekreten* ('strictly confidential') vs. the synonymous expression *striktno sekreten*.

Idiomaticity is a very broad concept. Here we use the term mainly with respect to the restrictions idiomaticity imposes on the morphosyntactic form, the semantics and the statistical frequency of MWEs. The degree of idiomaticity, or markedness, determines the way MWEs are treated in various NLP applications, such as, for example, Machine Translation. Compositionality represents the degree to which the complex meaning of the MWE is a combination of its components. After a MWE is formed, it enters into paradigmatic and syntagmatic relations in the lexical system. This means that in its context of use a MWE may change its compositionality and respectively – its level of idiomaticity.

For example, the phrase *poshtenska kutiya* ('post box') is formed as a regular decomposable combination where the adjective *post* (relating to a postal service) and *box* are realised with their usual lexical meanings. In recent years the phrase acquired an additional meaning – 'electronic post box, email', which is clearly idiomatic although the origins of the phrase and the relation between the components is still easily recoverable.

2.2. Classification of MWEs with respect to Idiomaticity

We adopt the general classification of MWEs presented by Baldwin et al. (2003). The authors distinguish between the following three categories: (a) non-decomposable MWEs for which a decompositional analysis of the meaning is not possible, e.g. *shepherd's purse*; (b) idiosyncratically decomposable MWEs for which some components of the phrase have a meaning not observed independently outside the MWE, e.g. *periodic table*; and (c) simple decomposable MWEs whose meaning can be decomposed to that of their constituents but nonetheless comprise a single lexical unit, e.g. *Bulgarian language*. For instance, due to institutionalisation simple decomposable MWEs often exhibit restrictions in the syntactic structure or synonym substitutions within the MWE. In these respects they differ from free phrases which are decomposable and are not considered lexical units, e.g. *important factor*.

For the purposes of some applications we may be interested simply in distinguishing between MWEs and free phrases in order to define separate methodologies for their treatment, e.g. keyword extraction,

while in other cases a more detailed categorisation may be required because the categories of MWEs differ with respect to their characteristic features and thus pose different problems, e.g. Machine Translation. On the one hand, the non-decomposable MWEs need to be defined in a dictionary so that they can be supplied with suitable translations. On the other hand, it is inefficient to add decomposable MWEs to the dictionary as their number is large and their meaning is defined as a function of their constituents. Different translation approaches may be adopted depending on the features of the different types of MWEs. Therefore, in many cases we are interested not only in recognising MWEs but also in discriminating between different categories of MWEs.

We divide simple decomposable MWEs into ten categories based on the following semantic and pragmatic factors: (1) Reference to NEs: (i) whether they contain a NE; and/or (ii) whether they constitute a NE; (2) Degree to which the connection between the components is explicit or can be restored. The classification is based on idiomaticity (Stoyanova, 2012):

(1) NEs without an (evident) connection between the elements – e.g., personal names *Ivan Petrov*. These are more often transliterated into other languages rather than translated, unless there is an established form for the NE in the target language.

(2) NEs with a meaningful element – e.g., *Stara Zagora* (literally, 'Old Zagora'), *North Korea*. The meaningful component is very often translated.

(3) Non-NEs with a vague connection between the components – e.g., *cave lion*. Most often these MWEs cannot be translated literally but have an established equivalent, e.g. *vodno konche* (literally, 'water horse') whose equivalent in English is 'dragonfly'.

(4) NEs containing meaningful components with difficult to restore connection – e.g., *Black Sea*. The approach to their rendition in other languages is mixed – some components may be translated and others transliterated, depending on how much of the linking information can be restored.

(5) NEs consisting of a descriptor and a NE, e.g. *Treaty of London*. These MWEs are usually translated, often rendered literally. Even if the translation of the NE is not fully equivalent to the original in meaning, the NE is still recognisable.

(6) Non-NEs which contain a NE as one of its components – *Down syndrome*. Similar to (5).

(7) Non-NEs with a standard, easy to restore connection between the components, e.g. *sea turtle* where the connection between the components is 'habitat' – 'turtle inhabiting the sea'. Categories 7-10 are very often translated literally since these are mostly descriptive decomposable MWEs.

(8) NEs with a standard, easy to restore connection between the components – *Association for Computational Linguistics*.

(9) Non-NEs with an explicit connection between the components – *self-retracting knife*. There is a subtle difference between categories (7) and (9) – in the latter the connection is explicit (e.g. 'retracts itself'), while in (7) it is not present in the MWE but is easy to recover (e.g. 'sea' is habitat). Same correspondence exists between categories (8) and (10). Explicit connection usually implies the presence of a verbal component – a participle or a verbal adjective or noun. The corresponding categories with explicit/easy to recover connection usually receive simillar treatment in automatic processing.

(10) NEs with an explicit connection between the components – *Center for the Treatment and Study of Anxiety*.

(*) Free collocations – *chist vazduh* ('fresh air'). Free collocations are free phrases (non-MWEs) which are statistically marked, i.e. they appear with high frequency compared to other synonymous candidates but are not linguistically (lexically, semantically or morphosyntactically) marked. Here they are included for completeness.

On the one hand, NEs are strongly institutionalised, which means that they may have an established translation different from the literal one, and the translation variants might be restricted. For example, the NE *Organizatsiya na obedinenite natsii* (literally 'Organisation of the United Nations') in Bulgarian differs from its English correspondence *United Nations*. On the other hand, MWEs which are not NEs are usually less restricted and allow certain variations.

The composition of MWEs often imposes different restrictions mainly on the subordinate components. Firstly, these are grammatical constraints – agreement between the subordinate part and the head (A N phrases). Some cases, however, require additional restrictions on the subordinate component which can further be used for the successful identification of MWEs. Prepositional phrases in MWEs usually express a class of objects but not a concrete object, for example *pasta za zabi (toothpaste)* – literally, 'paste for teeth', is a MWE, while *pasta za zabite na Ivan* ('paste for Ivan's teeth') is not a MWE, **pasta za zab* ('paste for a tooth'), **pasta za zabite* ('paste for the teeth') are unacceptable (their frequency in BNC is 0 compared to 417 occurrences of *pasta za zabi* ('toothpaste').

The modifications of decomposable MWE components are not always strictly restricted as in the other categories of MWEs. Although the MWE denotes a single concept, in some cases component modifications are allowed which leads to concept modification and a different meaning. It may result in the composition of a new lexical item – for example, *pasta za mlechni zabi* ('toothpaste for milk teeth') considered as a separate MWE, hyponym of *pasta za zabi* ('toothpaste'), or of a free phrase where the meaning of a component is concrete – for example, *torta s morkovi* ('carrot cake') → *torta s morkovite ot gradinata* ('cake with the carrots from the garden').

3. Method for Automatic Categorisation of MWEs Based on Idiomaticity

The method presented here is focused on MWE categorisation for the purposes of automatic text processing of Bulgarian. Different types of MWEs exhibit distinctive features and thus require specific treatment with regards to various applications (see section 2.2.).

The method is applied on annotated Bulgarian texts – sentence splitting, POS tagging, grammatical characteristics. The type of nouns – common or proper, has also been assigned. The method comprises the following rules:

1. Given that a MWE consists only of words recognised as proper nouns, classify it as a NE (category 1).

2. Given that a MWE consists of a proper noun and other elements and all the words begin with a capital letter, classify it as category 2.

3. Given that a MWE consists of a proper noun and other words and the first word of the MWE begins with a capital letter, classify it with the greatest probability as category 4 or 5.

4. Given that a MWE includes a proper noun and the first word of the MWE does not begin with a capital letter, classify it with the greatest probability as category 6.

5. Given that a MWE does not include a proper noun and the MWE begins with a capital letter, classify it with the greatest probability as category 8 or 10.

6. Given that a MWE does not include a proper noun and does not begin with a capital letter, classify it with the greatest probability as category 3, 7 or 9.

Figure 1 sketches the algorithm used for automatic detection of the MWE categories on the basis of the proposed rules.

More fine-grained categorisation might be achieved if we introduce some more specific rules incorporating semantic analysis such as Latent Semantic Analysis (Landauer et al., 2007), or lexical-semantic information from WordNet such as noun labels (e.g., noun.location) or semantic relations. In some cases it is sufficient to determine the group of categories the MWE belongs to, depending on the purposes of the study, and it may be inefficient to unambiguously assign a single category.

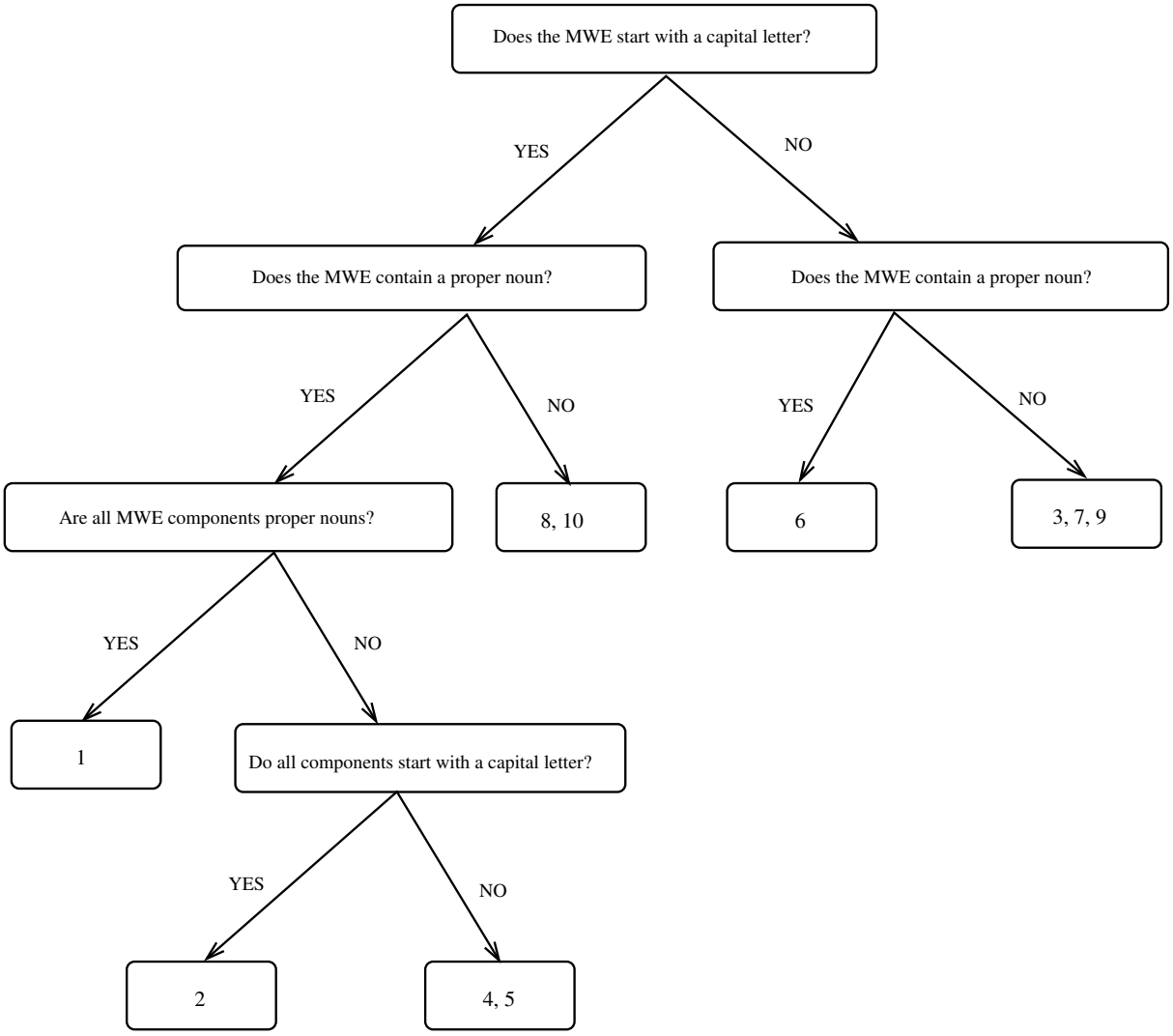

Figure 1: Algorithm for MWE category recognition.

The most problematic is the combined group of categories 3, on the one hand, and 7 and 9 on the other, since these can share the same form but have different semantic structure and thus may require different processing and analysis. Moreover, there are many MWEs which are on the boundary between categories and it can be difficult to distinguish between them.

It should be mentioned that the rules rely on some language-specific information such as the use of capital letters. However, there are many other Slavic and European languages which share these rules – capitalising first letter of names; common nouns are not capitalised (or only a limited numbers of categories are – months, days of the week); etc. The rules in this form have limited applicablity for German and other languages which capitalise all nouns, although they can be adapted and/or extended accordingly.

4. Experiment

4.1. Linguistic Resources

The experiments are based on the `Wiki1000+` corpus which comprises 6,311 Wikipedia articles, each of them containing at least 1000 words. The corpus amounts to 13.4 million words of running text distributed between 25 domains (Leseva and Stoyanova, 2014). The corpus has been supplied with linguistic annotation which includes several components – sentence segmentation, tokenisation, POS tagging and lemmatisation. The annotation is performed automatically using the set of tools of the Bulgarian

Language Processing Chain (Koeva and Genov, 2011). The POS tagger also assigns additional lexical information including the type of each noun – common or proper, and the grammatical characteristics of the word.

Syntactic type	# entries	% of all
(A) N	16,791	28.3
N N	35,314	59.5
N PP	4,424	7.5
(A) N PP	965	1.6
Other	1,875	3.1

Table 1: Syntactic types of MWEs in the dictionary (A=Adjective, N=noun, P=Preposition, PP=Prepositional phrase; brackets denote possible repetition, i.e. (A)N includes phrases of the form AN, AAN, etc.)

Idiomatic type	# entries	% of all
NE	39,982	67.3
non-NE	13,774	23.2
NE, contains-NE	3,339	5.6
non-NE, contains-NE	1,672	2.8
Unclassified	602	1.0

Table 2: Idiomatic types of MWEs in the dictionary

Additionally, noun phrases (NPs) in `Wiki1000+` have been identified using a list of possible syntactic constructions, and all MWEs have been annotated by applying a large dictionary containing over 85,000 MWEs, of which 59,369 NPs (Todorova and Stoyanova, 2014). The distribution of dictionary entries in terms of their syntactic structure is presented in Table 1, while their distribution with respect to references to NEs is shown in Table 2. Table 3 shows the result of the annotation of the different MWE categories in the corpus using the MWE dictionary.

Category	Label	#	% of all MWE
Non-decomposable	A	700	0.23
Idiosyncratically decomposable	B	3,156	1.02
Category	1	36,932	11.95
Category	2	11,248	3.64
Category	3	1,461	0.47
Category	4	1,086	0.35
Category	5	18,962	6.13
Category	6	27,373	8.86
Category	7	140,394	45.42
Category	8	16,653	5.39
Category	9	1,468	0.47
Category	10	0	0
”Free collocations”	X	49,651	16.06
Free phrases	Y	1,197,762	-

Table 3: Distribution of types of MWEs in Wiki1000+ corpus

The corpus `Wiki1000+` and the MWE dictionary are distributed as part of META-SHARE[1].

[1] `http://metashare.ibl.bas.bg/repository/search/`

4.2. Tasks

In order to observe the performance of the method, two distinct sets of tasks were defined.

1. Automatic MWE categorisation without prior MWE recognition – in this case the method for categorisation is applied on all NPs. It involves the following steps:
   - POS tagging and lemmatisation;
   - identification of NPs and syntactic filtering;
   - categorisation on all identified NPs.
2. Automatic MWE categorisation following MWE recognition – in this case categorisation is applied only on NPs identified as MWEs. It includes:
   - POS tagging and lemmatisation;
   - identification of NPs and syntactic filtering;
   - identification of MWEs;
   - categorisation of recognised MWEs and identification of certain types of NEs.

The MWE categorisation method is applied independently of MWE recognition although they generally complement each other. The MWE recognition method used in the experiments is outlined below, but it falls outside of the scope of the present work. The experiments are limited to several NP constructions: (A) N; N N; N P N; and N P (A) N.

The method for MWE identification combines collocation extraction with syntactic filtering to eliminate invalid or rare constructions. The method is described by Justeson and Katz (1995). It gives relatively good results taking into account its simplicity and the limited resources it requires (only POS annotation is needed). However, this method is best suited for extracting MWEs with adjacent components and additional processing is required to adapt it for the task of identifying non-adjacent MWEs.

In our application of the method, mutual information (MI) is adopted as the associaton measure used for deciding whether the cooccurring words form a collocation (Manning and Schutze, 1999). Other measures have also been experimented with, such as the Chi-square, Log-likelihood, Dice coefficient, but they have not proven to be empirically superior to MI for our data. It is recognised that MI, as well as most of the other statistical measures, does not work well for low frequency events so we only consider N-grams with frequency of over 10 occurrences.

In order to evaluate the performance of the MWE categorisation method (in the second set of tasks) independently of the quality of MWE recognition, we perform the method on automatically annotated and manually verified MWEs from `Wiki1000+`. However, it should be noted that in real-life applications MWE categorisation is interweaved with MWE recognition and thus the performance of the categorisation is influenced by the results of the recognition.

The two sets of tasks are evaluated independently in order to establish whether MWE categorisation can be used for MWE identification as well. The nature of the rules suggested that the method can be applied with relative independence for the identification of some categories of NEs, although it is not suitable for non-NE MWE identification in general.

4.3. Results

Table 4 presents the results for different MWE categories in terms of precision and recall. The simple rule-based approach on already recognised MWEs reaches precision of 91.51% with variation of $\pm$ 4% (except category 6, see Table 4), while on unlabelled NPs the precision varies considerably between categories and ranges between 25.11% and 81.43%. Even for the categories with best results (category 1, with the vast majority of entities being personal names) the precision without prior MWE recognition is considerably lower (81.43%) than the precision after MWE recognition (94.10%) although the recall is slightly better.

The results confirm the hypothesis that the method is unsuited for MWE recognition on its own and does not obtain satisfactory results when applied independently on general NPs.

Category	Label	Precision	Recall
Non-decomposable and Idiosyncratically decomposable	A and B	77.5	82.9
Category	1	94.1	96.5
Category	2	89.0	91.7
Category	3	0	0
Category	4	0	0
Category	5	89.4	71.0
Category	6	79.6	90.8
Category	7	90.1	87.2
Category	8	87.4	87.3
Category	9	0	0

Table 4: Results (precision and recall) for different categories after MWE recognition

Categories (3) and (9) are grouped with category (7), and category (4) is grouped with (5), they are not recognised separately, therefore they appear with zero precision and recall in Table 4. In the application of the method after MWE recognition, errors are mainly due to combination of categories, errors in tagging, or specific cases of capital letter use. For more precise results it is required to pose additional constraints on the rules or involve more detailed structural and semantic information. Moreover, improving MWE recognition methods will invariably lead to improvement in MWE categorisation.

5. Related Work

Research in the field of automatic MWE recognition and analysis in the last few decades has been clearly divided into two main trends – on the one hand, unsupervised resource-light highly efficient but less effective statistical approaches, and on the other hand, linguistically based resource-dependent but often inefficient methods. Recent research suggests that successful MWE recognition and tagging lies in the balanced hybrid approaches.

The detailed linguistically motivated characteristic of MWEs both as morpho-syntactic and semantic units, is a necessary prerequisite for successful automatic rendition. In this respect our research relies on the theoretical and applied studies focused on MWE classification by Baldwin et al. (2003), Baldwin (2004), Nunberg et al. (1994), Sag et al. (2002), among others.

Hybrid methods for MWE identification have been applied and described by Justeson and Katz (1995), Smadja (1993), Baldwin et al. (2003), Widdows (2008), Nakov (2008), Giesbrecht (2009) and many others. The specific problems of the description and automatic recognition of MWEs and NEs in Bulgarian have been discussed by Koeva (2006), Koeva (2007), Todorova (2006), Todorova and Obreshkov (2008), Leseva and Stoyanova (2008).

6. Conclusion

In conclusion, the methods described in the paper are relatively simple and do not require elaborate linguistic resources. Thus, they are suitable for morphologically rich languages, such as Bulgarian.

We need to emphasize that the results presented here are only valid for noun phrases of a limited variety of syntactic structures, and the possible generalisation of the observations over the whole group of MWEs is still to be evaluated.

However, we can conclude that the approach described here can potentially be developed into a successful methodology by considering the parameters of the particular research purpose – whether we need to simply identify MWEs, or discriminate between categories, as well as the granularity of the categorisation. It is also important to consider the characteristics of the resources as they influence highly the results, and take into account the specific features of the analysed corpora and the employed dictionaries in the anlysis and evaluation. The extensive application and testing of methods for MWE identification remains one of the major tasks in natural language processing of Bulgarian.

References

Baldwin, T., Bannard, C., Tanaka, T., and Widdows, D. (2003). An Empirical Model of Multiword Expression Decomposability. In *Proceedings of the ACL Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 89–96.

Baldwin, T. (2004). *Multiword Expressions, Advanced course at the Australasian Language Technology Summer School (ALTSS)*.

Baldwin, T. (2006). *Compositionality and Multiword Expressions: Six of One, Half a Dozen of the Other?, COLING/ACL 2006 Workshop on MWEs*.

Giesbrecht, E. (2009). In Search of Semantic Compositionality in Vector Spaces. In Rudolph, S., Dau, F., and Kuznetsov, F. O. (eds.), *ICCS 2009, LNAI*, volume 5662, page 173–184. Springer-Verlag Berlin Heidelberg.

Justeson, J. and Katz, S. (1995). Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text. *Natural Language Engineering*, pages 9–27.

Koeva, S. and Genov, A. (2011). Bulgarian Language Processing Chain. In *Proceedings of Integration of Multilingual Resources and Tools in Web Applications. Workshop in conjunction with GSCL 2011, 26 September 2011, University of Hamburg*.

Koeva, S. (2006). Inflection Morphology of Bulgarian Multiword Expressions. In *Computer Applications in Slavic Studies*, pages 201–216. Boyan Penev Publishing Centre, Sofia.

Koeva, S. (2007). Multi-word Term Extraction for Bulgarian. In *Balto-Slavonic Natural Language Processing 2007*, pages 59–66, Prague.

Landauer, T. K., McNamara, D., Dennis, S., and Kintsch, W. (eds.). (2007). *Handbook of Latent Semantic Analysis*. Mahwah NJ: Lawrence Erlbaum Associates.

Leseva, S. and Stoyanova, I. (2008). Treatment of Named Entities in Machine Translation. In Blanko, X. and Silberztein, M. (eds.), *Proceedings of the 2007 International Nooj Conference*, pages 254–272, Barcelona, Spain. Cambridge Scholars Publishing.

Leseva, S. and Stoyanova, I. (2014). Wikipedia as a Source for Linguistic Resources – Corpora, Dictionaries and Language Models. In Koeva et al. (ed.), *Language Resources and Technologies for Bulgarian (in Bulgarian)*. In press.

Manning, C. and Schutze, H. (1999). *Foundations of Statistical NLP*. MIT Press.

Nakov, P. (2008). Noun Compound Interpretation Using Paraphrasing Verbs : Feasibility Study. In *Proceedings of the 13th International Conference on Artificial Intelligence Methodology Systems Applications AIMSA08 (2008)*, pages 103–117.

Nunberg, G., Sag, I. A., and Wasow, T. (1994). Idioms. *Language*, 70:491–538.

Sag, I., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*, Mexico.

Smadja, F. (1993). Retrieving Collocations from Text: Xtract. *Computational Linguistics*, 19:143–177.

Stoyanova, I. (2012). *Automatic Recognition and Tagging of Compound Lexical Units in Bulgarian (PhD thesis, in Bulgarian)*.

Todorova, M. and Obreshkov, N. (2008). Compilation of Inflectional Dictionaries Using WordEditor. In Tadić, M., Koeva, S., and Dimitrova-Vulchanova, M. (eds.), *Proceedings of the 6th International Conference on Formal Approaches to South Slavic and Balkan Languages*, pages 131–138.

Todorova, M. and Stoyanova, I. (2014). MWE Dictionaries in Bulgarian. In Koeva et al. (ed.), *Language Resources and Technologies for Bulgarian (in Bulgarian)*. In press.

Todorova, M. (2006). On The classification of Bulgarian Non-Free Phrases. In Vulchanova, M. D., Koeva, S., Krapova, I., and Vulchanov, V. (eds.), *Proceedings of the Fifth International Conference Formal Approaches to South Slavic and Balkan Languages*, pages 251–256, Sofia, Bulgaria.

Widdows, D. (2008). Semantic Vector Products: Some Initial Investigations. In *Proceedings of the Second AAAI Symposium on Quantum Interaction*.