



# Identification of Metaphorical Collocations in Different Languages – Similarities and Differences

Lucia Nacinovic Prskalo<sup>(✉)</sup>  and Marija Brkic Bakaric 

Faculty of Informatics and Digital Technologies, University of Rijeka, Radmile  
Matejcic 2, 51000 Rijeka, Croatia  
{lnacinovic,mbrkic}@uniri.hr

**Abstract.** Metaphorical collocations are a subset of collocations in which a semantic shift has occurred in one of the components. The main goal of this paper is to describe the process of identifying metaphorical collocations in different languages – English, German and Croatian. Approaches to annotating metaphorical collocations from a list of word sketches for the three languages are presented using one of the most common nouns for all three languages – “year” for English, “Jahr” (Engl. year) for German, and “godina” (Engl. year) for Croatian. The compilation of a list of relevant grammatical relations in the identification of metaphorical collocations for each language is also described. Finally, the procedures for automatic classification of metaphorical collocations for Croatian, German and English are performed and compared.

**Keywords:** Metaphorical collocations · Collocations · Classification · Automatic collocation identification · Grammatical relations

## 1 Introduction

Metaphorical collocations are a subset of collocations in which there is a shift in the meaning of a component, usually the collocate. In this paper, procedures for identifying metaphorical collocations are described and compared for three different languages - English, German and Croatian. Procedures for (automatic) identification of metaphorical collocations require a number of sub-procedures, such as corpus selection, creation of an inventory of collocations and metaphorical collocations, selection of relevant grammatical relations, machine learning modelling, etc. They are described in details in the following sections.

According to research by [1], it was found that the noun mostly appears in the role of the base, while verbs and adjectives, which have the potential to acquire secondary meaning leading to polysemy, mostly appear in the role of the

---

Supported by Croatian Science Foundation under the project *Metaphorical collocations – Syntagmatic word combinations between semantics and pragmatics* (IP-2020-02-6319).

collocate. In the same article, the authors give an example of the metaphorical collocation in Croatian “zabiti gol” (Engl. to kick a goal), where the noun “gol” is the base and the verb “zabiti” has the function of a collocate. In the same example, the idiosyncratic character of collocational compounds, in which two lexemes are arbitrarily connected in a collocation, is evident. Therefore, it is very difficult for a non-native speaker to know which word combinations are common in another language and which are not acceptable. For example, the metaphorical collocation “zabiti gol” - (Engl. to hit a goal) mentioned above uses different collocates in other languages - in German the equivalent translation is “ein Tor schießen” - (Engl. to shoot a goal), and in English “kick a goal”. The authors note that the collocate is expressed in different images in these languages and that the meaning is determined by different extralinguistic comparisons. However, they also note that the comparisons are based on the same concept, namely the physical encounter of the ball and the goal. They conclude, therefore, that the examples of collocation compounds in different languages indicate that the same extra-linguistic reality is lexicalized in different ways, which superficially gives the appearance of arbitrariness, but that the process of forming collocation compounds seems to follow the same pattern. Testing this assumption is one of the main long-term goals of our project, and the research presented in this paper represents the first steps in this direction.

Throughout our project, we intend to use a combination of computational-linguistic and theoretical-semantic approaches to obtain the most relevant results as quickly as possible. Manual annotation of metaphorical collocations in the corpus is an extremely time-consuming and tedious task. Therefore, by combining the two approaches, we aim to facilitate the task of finding different types of metaphorical collocations in different languages and identifying similarities and differences in the formation of metaphorical collocations in different languages.

In the second section we have singled out some of the important related work. The third section describes the research methodology - corpus selection, selection of the most relevant grammatical relations, annotation procedures, and model development using machine learning techniques. Finally, a description of the results and a conclusion follow.

## 2 Related Work

There is research that has addressed the automatic identification of metaphors in a text, such as in [2–5], but to our knowledge there is no research that has addressed the extraction of metaphorical collocations as a subset of collocations in general. The work in the remainder of this section relates to the extraction of collocations in general.

Most of the papers on the topic of automatic collocation extraction have dealt with the application of various association measures. For example, Church and Hanks [6] proposed an association ratio measure in 1990, Dunning [7] proposed a likelihood ratio measure in 1993, and Kita [8] and co-workers proposed a cost criteria measure in 1994. Smadja et al. [9] use the Dice coefficient to find translations of a source language collocation using a parallel corpus, and Thanopoulos

et al. [10] propose a PMI (Pointwise mutual information) measure. All these works focus on statistical methods in the form of association measures.

Some papers have experimented with parsing in addition to statistical measures. For example, Seretan and Wehrli [11] use different types of parsing for flexible pair extraction, and Lin [12] proposes an approach using a parser to extract dependency triples from a corpus.

There have also been attempts to include other linguistic features in the models, such as POS tags. For example, Krenn [13] uses statistical POS taggers and a partial parser to extract collocation-specific syntactic constraints. Pearce [14] uses WordNet to present an approach based on constraints on the possible substitutions of synonyms within candidate phrases.

Karan et al. [15] evaluate classification algorithms and features for the purpose of collocation extraction in Croatian. They use several classification algorithms and conclude that the SVM classifier performs best on bigrams and the decision trees on trigrams.

Ljubešić et al. [16] compare two approaches to the ranking of collocates - the logDice method and the word embedding method - on two Slovenian data sets. The quantitative evaluation shows that the machine-learning approach gives better results, but the lexicographers consider the obtained listings of collocates to be very similar.

A detailed systematic literature review on collocation extraction can be found in [17].

### 3 Methodology

One of the main goals of our research is to extract metaphorical collocations in Croatian, English, German and Italian. We also want to find out if there are similarities and peculiarities in the creation and extraction of metaphorical collocations in different languages.

We are currently in the first phase of research, in which we are detecting the basic metaphorical collocations and investigating their composition. The second phase will involve the detection of translation equivalents and their extraction.

There are no evaluation resources for extracting metaphorical collocations. In order to evaluate our own experiments, we are forced to develop suitable gold standard reference data sets ourselves.

The following subsections describe the main steps we have taken so far. First, we describe the selected corpora. Then we give an overview of the grammatical relations of Croatian, German and English as used in the Sketch Engine [18] tool for creating the lists of word sketches.

The study begins with the noun as a part of speech, since it was determined to be the most productive part of speech in terms of creating metaphorical collocations. Therefore, the most frequent nouns in the Croatian language were identified. We used the Croatian Web Corpus [19] for that purpose. Since nouns that are not productive in collocation matching (e.g. proper nouns) also appeared among the identified nouns, such nouns were additionally excluded manually.

The same procedure was performed for the comparable corpora enTenTen20 for English [20], deTenTen18 [20] for German, and itTenTen20 [20] for Italian. The nouns overlapping in four identical corpora were selected to ensure empirically relevant data. Then, the collocation profiles of the most frequently occurring noun “year” are analyzed in Croatian (“godina”), English and German (“Jahr”), and the collocation compounds expressing a metaphorical meaning are manually annotated.

### 3.1 Corpora

As mentioned earlier, for the Croatian language, we used the Croatian Web Corpus [19], which consists of texts collected on the Internet and contains over 1.2 billion words. The hrWaC corpus was PoS-tagged using the MULTEXT-East Croatian POS tag-set version 5 [21].

The English Web Corpus (enTenTen20) [20] is an English language corpus consisting of texts collected from the Internet. The latest version of the enTenTen corpus contains more than 38 billion words. The authors state that sample texts from the largest web domains, which account for 40% of all texts in the corpus, were manually reviewed and content with poor text quality and spam was removed. The corpus was tagged with the TreeTagger tool using the English Web 2020 part-of-speech tag set.

The German Web Corpus (deTenTen18) [20] is a German corpus that also consists of texts collected from the Internet. The latest version includes 5.3 billion words. The corpus contains lemmatization including gender lemmas and part-of-speech tagging. It was annotated with the RFTagger tool using the German RFTagger part-of-speech tag-set.

### 3.2 Grammatical Relations and Annotation

When we use the Word Sketch function in the Sketch Engine, we get a list of word’s collocates and other words in their environment, organized into categories called grammatical relations. These are, for example, words that serve as the subject of the verb, words that modify the word, etc. Each language/corpus has differently defined grammatical relations, which are specified by rules in the Sketch grammar. Sketch Engine uses the logDice measure [22] to identify collocations.

By default, the word sketches are sorted so that the sketches with the highest logDice score come first. logDice can be thought of as “typicality”. A high score means that the collocate is frequently found with the base and there are not many other bases with which the collocate is combined. In this case, it is a strong collocation. A low score means that the collocate tends to combine not only with that particular base, but with many other words. In this case, it is a weak collocation.

The linguists analyzed the lists thus obtained for the most frequent word in all three languages (Croatian, German and English) - “godina”, “Jahr” and “year” - and performed the annotation of collocations and metaphorical collocations.

Before compiling the final list of metaphorical collocations for each language, the experts held several discussion sessions until they felt confident enough to distinguish between the different types of collocations and thus extract metaphorical collocations. Two linguists per language participated in the task. The annotation varies in detail among the languages, so we provide the details for each language below.

The preliminary results of the analysis show that in all three languages, the most productive grammatical relations and combinations of parts of speech with respect to metaphorical collocations are combinations of an adjective in the function of a collocate + a noun in the function of the base and a verb in the function of the collocate + a noun in the function of the base. It is also noticed that phrasal verbs in English make a productive relation in terms of forming metaphorical collocations.

**Croatian.** For the word “godina”, the Sketch Engine generates a total of 21 grammatical relations: *kakav?*, *oba-u-genitivu*, *u-genitivu-n*, *a-koga-čega*, *n-koga-čega*, *koga-što*, *particip*, *prijedlog*, *infinitive*, *koga-čega*, *s-prilogom*, *a-koga-što*, *a-komu-čemu*, *komu-čemu*, *glagol-ispred-prijedloga*, *prijedlog-iza*, *veznik*, *koordinacija*, *imenica-iza-prijedloga*, *biti-kakav?* and *subjek-od*. There are 1,747 unique collocates dispersed over different grammatical relations out of a total of 5,019 collocation candidates. After extensive analysis by linguists, it was decided that the following grammatical relations are most productive and relevant for the formation and identification of metaphorical collocations in Croatian: **kakav?** (like what), **n-koga-čega** (two nouns - one in genitive), **koga-što** (accusative), **subjekt-od** (subject of), **particip** (participle), **biti-kakav?** (be like what). Detailed explanations on why those relations were chosen can be found in [1] and [23]. The annotated data set for Croatian also includes labels for the type of metaphorical collocation, such as for example “lexicalized metaphor”, “metaphor”, “term-metonymy”, etc.

The statistics of the annotated data set for Croatian is shown in Table 1.

**Table 1.** Statistics of the annotated data set for Croatian.

Relation	# of cands	# of colls	# of m-colls	Ratio of m-colls
kakav?	99	54	54	55%
n-koga-čega	100	41	38	41%
koga-što	100	41	41	41%
particip	100	16	11	11%
subjekt-od	100	30	30	30%
biti-kakav?	74	20	20	55%
<b>Total</b>	<b>673</b>	<b>202</b>	<b>193</b>	<b>29%</b>

**German.** For the word “Jahr”, the Sketch Engine generates a total of 18 grammatical relations: *Constructions, modifiers of Jahr, verbs with Jahr as subject, genitive objects of Jahr, nouns with Jahr as genitive object, dative objects of Jahr, nouns with Jahr as dative object, accusative objects of Jahr, nouns with Jahr as accusative object, verbs with Jahr as genitive object, verbs with Jahr as dative object, verbs with Jahr as accusative object, Jahr and/or ..., prepositions with Jahr as object, prepositional phrases, prepositional objects in dative, prepositional objects in accusative, prepositional objects in genitive.* After extensive analysis by linguists, it was decided that the following grammatical relations are most productive and relevant for the formation and identification of metaphorical collocations in German: **modifiers of x** (e.g. kommende Jahr), **verbs with x as subject** (e.g. das Jahr beginnt), **verbs with x as accusative object** (e.g. Jahr verbringen) and **nouns with x as genitive object** (e.g. Anfang des Jahres). Similar to the Croatian dataset, the annotated data set for German also includes labels for the type of metaphorical collocation, such as “lexicalized metaphor”, “metaphor”, “term-metonymy”, etc.

The statistics of the annotated data set for German is shown in Table 2.

**Table 2.** Statistics of the annotated dataset for German.

Relation	# of candS	# of collS	# of m-collS	Ratio of m-collS
Modifier of x	105	57	41	39%
Verbs with x as subject	100	86	13	13%
Verbs with x as accusative object	101	66	33	33%
Nouns with x as genitive object	403	349	51	13%
<b>Total</b>	<b>709</b>	<b>558</b>	<b>138</b>	<b>19%</b>

**English.** For the word “year”, the Sketch Engine generates a total of 27 grammatical relations out of which 14 are of the type verbs with particle “x” and “year” as object: *modifiers of “year”, nouns modified by “year”, verbs with “year” as object, verbs with “year” as subject, “year” and/or ..., prepositional phrases, adjective predicates of “year”, “year” is a ..., year’s ..., possessors of “year”, pronominal possessors of “year”, ... is a “year”, verbs with particle “x” and “year” as object, usage patterns.* After extensive analysis by linguists, it was decided that the following grammatical relations are most productive and relevant for the formation and identification of metaphorical collocations in English: **modifiers of “year”, verbs with “year” as object, verbs with “year” as subject, and verbs with particle “x” and “year” as object.** For English, there are no annotations yet for the type of metaphorical collocation. There are also no annotated collocations for a part of the list, but only metaphorical collocations are annotated.

The statistics of the annotated data set for Croatian can be found in Table 3.

**Table 3.** Statistics of the annotated data set for English.

Relation	# of candS	# of m-colls	Ratio of m-colls
modifiers of “year”	94	28	30%
verbs with “year” as object	98	13	13%
verbs with “year” as subject	100	7	7%
verbs with particle “x” and “year” as object	541	109	20%
<b>Total</b>	<b>833</b>	<b>157</b>	<b>19%</b>

### 3.3 Experiment

We experimented with models for automatic identification of metaphorical collocations from the lists created by the linguists described in the previous section. We trained the models for Croatian German and English, as they are fully completed at the time of writing.

**Preprocessing.** Before the model could be trained, all non-standard language variants, misspelled words, incorrectly lemmatized forms, duplicated lemmas, etc. had to be removed or corrected from the list of collocations and metaphorical collocations.

**Models.** We experimented with several models, including Support Vector Machines (SVM), Multilayer Perceptron, C4.5, and Random Forest. When training the models, we performed 10-fold cross-validation.

**Features.** As features for model training, we used collocation frequency, logDice, grammatical relation, and pretrained word embeddings (containing 300 vectors for each word) as implemented in fastText [24]. While frequency and logDice are statistical measures, grammatical relation represents syntactic information and word embeddings represent semantic information as they are vector representations of the context in which a word occurs.

## 4 Results

We evaluated the classification models based on Accuracy (percentage of correctly classified instances), Precision (proportion of correctly classified positive instances out of all positive instances in the system output), Recall (proportion of correctly identified positive instances out of all instances that should have been identified as positive) and F-measure (a measure that combines the results of Precision and Recall).

Table 4 shows the results for the Croatian data set, Table 5 the results for the German data set, and Table 6 for the English data set.

**Table 4.** Results for Croatian data set.

Model	Accuracy	Precision	Recall	F-measure
SVM	71.4706 %	0.715	0.715	0.714
Multilayer Perceptron	<b>75.2941 %</b>	<b>0.754</b>	<b>0.753</b>	<b>0.752</b>
C4.5	69.7059 %	0.697	0.697	0.697
Random Forest	68.8235 %	0.688	0.688	0.688

**Table 5.** Results for German data set.

Model	Accuracy	Precision	Recall	F-measure
SVM	<b>82.9023 %</b>	<b>0.865</b>	0.932	<b>0.897</b>
Multilayer Perceptron	79.454 %	0.853	0.898	0.875
C4.5	76.5805 %	0.855	0.853	0.854
Random Forest	81.3218 %	0.826	<b>0.971</b>	0.893

From the results in the tables, we can see that all the models used perform similarly within each language, with the best results obtained with the Multilayer Perceptron model for the Croatian data set (Acc 75.2941%, P 0.754, R 0.753, F 0.752), with the SVM model for the German data set (Acc 82.9023%, P 0.865, F0.897) and with the Random Forest model for the English data set (Acc 72.242%, R 0.911, F0.826). We obtained the best Recall (0.865) for the German data set with Random forest model and the best Precision (0.788) for the English data set with the C4.5 model. Comparing the results between the languages, the best results are obtained for German.

We also found that the inclusion of word embeddings in the features improved Accuracy by almost 10%.

**Table 6.** Results for English data set.

Model	Accuracy	Precision	Recall	F-measure
SVM	67.6157 %	0.743	0.842	0.789
Multilayer Perceptron	67.9715 %	0.752	0.830	0.789
C4.5	69.9288 %	<b>0.788</b>	0.798	0.793
Random Forest	<b>72.242 %</b>	0.755	<b>0.911</b>	<b>0.826</b>



## 5 Conclusion

In this paper, we have described the steps that we and our colleagues on the project team have taken to identify metaphorical collocations as a subset of the general category of collocations in different languages. So far, we have set the theoretical and methodological framework and taken the first steps to create the golden standard for Croatian, English and German. We have also determined the set of nouns we will use in our study and analyzed the word sketches for the most frequent noun in all three corpora - Croatian (“godina”), German (“Jahr”) and English (“year”).

The project team linguists also selected the most productive grammatical relations for the formation of metaphorical collocations. For Croatian, these are: *kakav?* (like what), *n-koga-čega* (two nouns - one in the genitive), *koga-što* (accusative), *subject-od* (subject of), *particip* (participle), *biti-kakav?* (to be like what). For German these are: modifiers of *x*, verbs with *x* as subject, verbs with *x* as accusative object, and nouns with *x* as genitive object. For English these are: modifiers of “year”, verbs with “year” as object, verbs with “year” as subject, and verbs with particles “*x*” and “year” as object.

The preliminary results of the analysis show that in all three languages the most productive grammatical relations and combinations of parts of speech in terms of metaphorical collocations are combinations of an adjective in the function of a collocate + a noun in the function of the base and a verb in the function of the collocate + a noun in the function of the base. Moreover, it is noted that phrasal verbs in English are a productive relation for the formation of metaphorical collocations.

The percentage of metaphorical collocations in the annotated data sets is slightly different for different languages - for Croatian it is 29%, for German 19% and for English 19%. It remains to be clarified whether these differences are due to different characteristics of the individual languages or to the different approaches of the annotators. It has already been noted in our project meetings that determining whether a collocation is also a metaphorical collocation might be subject to different approaches, since there are different types of metaphorical collocations (e.g. lexicalized metaphorical collocations, term, metonymy, etc.). It was also noted that further new guidelines need to be found to specify the annotation process for all languages as new insights are gained through the research process.

We also trained models for automatic recognition of metaphorical collocations from the candidate lists for Croatian, German and English created by the linguists. We experimented with four different models - Support Vector Machines (SVM), Multilayer Perceptron, C4.5, and Random Forest. We used collocation frequency, logDice, grammatical relation and pretrained word embeddings as features in model training. We obtained the best results with the Multilayer Perceptron model for the Croatian dataset (Acc 75.2941%, P 0.754, R 0.753, F 0.752), with the SVM model for the German data set (Acc 82.9023%, P 0.865, F0.897) and with the Random Forest model for the English data set (Acc 72.242%, R 0.911, F0.826). We obtained the best Recall (0.865) for the German data set

with Random forest model and the best Precision (0.788) for the English data set with the C4.5 model. We also found that the inclusion of word embeddings significantly improved the results. The results we obtained seem to be promising, but they can only be considered as preliminary results since they are based on only one most common noun. It remains to be seen how the data sets will behave for other nouns.

Our future work includes compiling similar lists and conducting experiments for other nouns. We also plan to test other measures and linguistic features to find methods that give the best results in extracting metaphorical collocations.

**Acknowledgement.** This work has been fully supported by Croatian Science Foundation under the project *Metaphorical collocations - Syntagmatic word combinations between semantics and pragmatics* (IP-2020-02-6319).

## References

1. Stojić, A., Košuta, N.: Izrada inventara metaforičkih kolokacija u hrvatskome jeziku - na primjeru imenice godina. *Fluminensia* **34**(1), 9–29 (2022)
2. Shutova, E.: Models of Metaphor in NLP. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 688–697. Association for Computational Linguistics, Uppsala, Sweden (2010)
3. Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., Dyer, C.: Metaphor detection with cross-lingual model transfer. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 248–258. Association for Computational Linguistics, Baltimore (2014)
4. Choi, M., Lee, S., Choi, E., Park, H., Lee, J., Lee, D., Lee, J.: MelBERT: metaphor detection via contextualized late interaction using metaphorical identification theories. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1763–1773. Association for Computational Linguistics (2021)
5. Wan, H., Lin, J., Du, J., Shen, D., Zhang, M.: Enhancing metaphor detection by gloss-based interpretations. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 1971–1981. Association for Computational Linguistics (2021)
6. Church, K.W., Patrick, H.: Word association norms, mutual information, and lexicography. *Comput. Linguist.* **16**(1), 22–29 (1990)
7. Dunning, T.: Accurate methods for the statistics of surprise and coincidence. *Comput. Linguist.* **19**(1), 61–74 (1993)
8. Kita, K., et al.: A comparative study of automatic extraction of collocations from corpora: mutual information vs. cost criteria. *J. Natural Lang. Process.*, 21–33 (1994)
9. Smadja, F.: Retrieving collocations from text: Xtract. *Comput. Linguist.* **19**(1), 143–78 (1993)
10. Thanopoulos, A., Fakotakis, N., Kokkinakis, G: Comparative evaluation of collocation extraction metrics. In: Proceedings of the Third International Conference on Language Resources and Evaluation, pp. 620–25 (2002)
11. Seretan, V., Wehrli, E.: Multilingual collocation extraction with a syntactic parse. *Lang. Resour. Eval.* **43**(1), 71–85 (2009)

12. Lin, D.: Extracting collocations from text corpora. In: Proceedings of the First Workshop on Computational Terminology, pp. 57–63 (1998)
13. Krenn, B.: Collocation Mining: Exploiting Corpora for Collocation Identification and Representation. *Entropy*, 1–6 (2000)
14. Pearce, D.: Synonymy in Collocation Extraction. In: Proceedings of the Workshop on WordNet and Other Lexical Resources, Second Meeting of the North American Chapter of the Association for Computational Linguistics, pp. 41–46. Association for Computational Linguistics, (2021)
15. Karan, M., Šnajder, J., Dalbelo Bašić, B.: Evaluation of Classification Algorithms and Features for Collocation Extraction in Croatian. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), pp. 657–62. European Language Resources Association (ELRA), Istanbul, Turkey (2012)
16. Ljubešić, N., Logar, N., Kosem, I.: Collocation ranking: frequency vs semantics. *Slovenscina 2.0*, **9**(2), 41–70 (2021)
17. Brkic Bakaric, M., Nacinovic Prskao, L., Matetic, M.: Insights into automatic extraction of metaphorical collocations. *Rasprave* (Manuscript submitted for publication)
18. Kilgarriff, A. et al.: The Sketch Engine: Ten Years On. *Lexicography*, pp. 7–36 (2014)
19. Ljubešić, N., Erjavec, T.: hrWaC and slWac: compiling web corpora for croatian and slovene. In: Habernal, I., Matoušek, V. (eds.) TSD 2011. LNCS (LNAI), vol. 6836, pp. 395–402. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-23538-2\\_50](https://doi.org/10.1007/978-3-642-23538-2_50)
20. Jakubiček, M. et al.: The TenTen Corpus Family. In: Proceedings of the 7th International Corpus Linguistics Conference CL, pp. 125–127 (2013)
21. Ljubešić, N. et al.: New inflectional lexicons and training corpora for improved morphosyntactic annotation of Croatian and Serbian. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA), Parise, France (2016)
22. Rychlý, P.: A Lexicographer-Friendly Association Score. In: Proceedings of the Recent Advances in Slavonic Natural Language Processing, pp. 6–9 (2008)
23. Stojić, A., Košuta, N.: METAPHORISCHE KOLLOKATIONEN - EINBLICKE IN EINE KORPUSBASIERTE STUDIE. *Linguistica* **61**(1), 81–91 (2022)
24. Grave, E., Bojanowski, P., Gupta, P., Joulin, A., Mikolov, T.: Learning Word Vectors for 157 Languages. In: Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), pp. 3483–3487 (2018). <https://fasttext.cc/>