A Multi-feature Classifier for Verbal Metaphor Identification in Russian Texts

Yulia Badryzlova[1(✉)] and Polina Panicheva[2]

[1] National Research University Higher School of Economics, Moscow, Russia
yuliya.badryzlova@gmail.com
[2] St. Petersburg State University, Saint Petersburg, Russia
ppolin86@gmail.com

Abstract. The paper presents a supervised machine learning experiment with multiple features for identification of sentences containing verbal metaphors in raw Russian text. We introduce the custom-created training dataset, describe the feature engineering techniques, and discuss the results. The following set of features is applied: distributional semantic features, lexical and morphosyntactic co-occurrence frequencies, flag words, quotation marks, and sentence length. We combine these features into models of varying complexity; the results of the experiment demonstrate that fairly simple models based on lexical, morphosyntactic and semantic features are able to produce competitive results.

Keywords: Sentence-level metaphor identification
Supervised binary classification · Feature engineering
Distributional semantic features · Lexical co-occurrence features
Morphosyntactic co-occurrence features

1 Introduction

Metaphor is said to be a ubiquitous yet a fugitive phenomenon: it resides in virtually every utterance of human language, but it is notoriously difficult to formalize. Not only is metaphor indispensable in various language processing tasks; it is also commonly accepted that metaphor is a pervasive process in human language and thought [20], with numerous effects in psychology, psycholinguistics, and cognitive disciplines.

Metaphor processing has attracted increasing attention and effort in recent years. A series of Workshops on Metaphor in NLP was held for several successive years as a part of the NAACL-HLT conference. The most comprehensive overview of approaches to automated metaphor identification is available in [41].

The following types of features are exploited in the state-of-the-art systems for metaphor identification in the supervised and the unsupervised settings:

- lexical [3, 4, 10, 15, 17, 22, 23, 28, 30];
- morphological [4, 15];
- distributional semantic [28, 31, 35, 38];
- topic modelling [4, 13];

D. Ustalov et al. (Eds.): AINL 2018, CCIS 930, pp. 23–34, 2018.
https://doi.org/10.1007/978-3-030-01204-5_3

- lexical thesauri and ontologies: WordNet [3, 10, 15, 17, 25, 27, 28, 33, 38, 39], FrameNet [11], VerbNet [3], ConceptNet [30], and the SUMO ontology [8, 9];
- psycholinguistic features [3, 10, 28, 29, 37–40];
- syntactic relations [15, 30].

Metaphor identification projects can be divided into two groups according to their theoretical premises. Experiments in the first group stem from the conceptual metaphor paradigm [20] which stipulates that linguistic metaphors are surface realizations of the underlying conceptual mappings between the source and the target domains. Projects of this type seek to identify evidence of such mappings in the text [e.g. 8–11, 13, 25, 27, 28, 30, 37]. The second vein of experimental research does not make any *a priori* assumptions about the underlying conceptual mechanisms of metaphor and searches for any stretches of metaphoric language in the text [e.g. 3, 4, 15, 17, 29, 31, 33, 35, 38–40].

Results of metaphor identification experiments are difficult to compare for a number of reasons: (a) the theoretical incompatibility and the subsequent differences in the experimental design; (b) some systems identify metaphors on the sentence level while others identify word-level metaphors; (c) many of the existing systems are domain-specific; and (d) most systems are trained and evaluated on different datasets.

Metaphor identification in Russian texts has been addressed in several projects. For example, [28, 30, 37] use a variety of features to model the conceptual source and target domains and to align them with their linguistic realizations in text, while [31, 38, 39] operate outside of the conceptual metaphor paradigm. The former two systems exploit cross-linguistic metaphors: the classifier is first trained on the English data, and then the trained model is projected to Russian using a dictionary. The latter project uses distributional semantic vectors to distinguish metaphoric and non-metaphoric sentences.

The subsequent sections of this paper describe a sentence-level Russian verbal metaphor identification experiment on raw text with a rich multi-feature classifier involving semantic, lexical, and morphological features, as well as information about the occurrence of flag words (specific lexical markers), quotation marks, and sentence length.

To the best of our knowledge, this is the first project outside of the conceptual metaphor paradigm to explore a model of such complexity for metaphor identification in Russian texts.

2 The Dataset

The experimental dataset is comprised of 7,166 sentences each of which contains one of the 20 polysemous Russian verbs (referred to as target verbs below); some of the experimental verbs are listed in Table 1. The full dataset and its description are available for download.[1]

[1] https://github.com/yubadryzlova/metaphor_dataset_20_verbs.git.

Table 1. Dataset: some of the target verbs

Russian	Transliteration	Translation (primary meaning)
бомбардировать	bombardirovat	to bombard (smth/smb))
доить	doit	to milk (e.g. a cow)
нападать	napadat	to attack (smth/smb)
отрубить	otrubit	to hack (smth) off
трубить	trubit	to blow a trumpet
уколоть	ukolot	to prick (smth/smb)
зажигать	zazhigat	to ignite (smth)

2.1 The Target Verbs

The verbs were chosen so as to match the specific linguistic properties:

– the verb has at least one primary meaning which is a typical meaning of Accomplishment or Activity [26, 42];
– the verb has at least one primary meaning which authorises a two-actant construction with the following mandatory actants: (1) the Agent, (2) the Patient/the Theme;
– the Agent denotes a human being(s); the other actants refer to physical (concrete, non-abstract) entities;
– the derivational structure of the verb's polysemy is transparent: each secondary meaning is derived from the primary one by means of either a metaphoric or a distant metonymic shift;
– the verb has a small number (<10) of meanings listed in the dictionary;
– the verb does not possess any strongly delexicalized meanings.

Verbs of this kind were chosen for the experiment because they bring the opposition of metaphoric and non-metaphoric meanings to its most distinct expression.

2.2 The Non-metaphoric and the Metaphoric Classes

The sentences in the dataset are divided into the two classes, the non-metaphoric and the metaphoric ones.

The Non-metaphoric Class. This class includes the sentences where the target verb is used either (a) in the central literal meaning (as described above) or (b) in the meanings that are related to the central meaning via either a diathetic shift (i.e. the change of the syntactic rank of the actants), or a close metonymic shift.

The Metaphoric Class. This class contains the three types of sentences: (c) conventionalized metaphors based on polysemy, (d) unconventional creative metaphors, and (e) idiomatic expressions.

Conventionalized Metaphors are the target verbs used in their secondary meanings. For example, consider the metaphoric meanings of *trubit* ‘to blow a trumpet’[2]:

- to talk profusely about smb, smth; to spread gossip, information, news, etc.;
- to perform a tiresome or tedious activity during a long period of time.

Unconventional Metaphors exploit the target verbs creatively to liken concepts from the target domain to concepts from the source domain [20] and to reinterpret the target in terms of the source, e.g. Сестра поглядела на нее, словно <уколола> кинжалом. ‘Sister threw a glance at her, as if she <pricked> her with a dagger.’

Idiomatic Expressions are fixed or semi-fixed compositional units whose meaning is not equal to the sum of the meanings of its constituent lexemes, e.g. Когда то мои пра - пра - пра - пра - прадеды … <грели руки> на ростовщичестве. ‘There was a time when my fore- fore-fore-fore-fore-forefathers used to <warm their hands> (= to make dishonest or illegal profit) with usury.’

Sentence Selection and Annotation The sentences were obtained from RuTenTen11, a 14.5 bn-word Russian web corpus, accessed via the SketchEngine interface [16]. The sentences were added to the dataset in the order in which they were retrieved, without any filtering. The selection of sentences and their annotation by the binary classes (metaphoric vs. non-metaphoric) was performed by one annotator, a trained linguist. The annotator was compelled to make binary decisions.

The subsets for the individual verbs are balanced by the class, i.e. 50% of the sentences are metaphoric while the other half are non-metaphoric. However, the dataset is not balanced across the verbs (ranging between 225 and 693 sentences per verb). The data is heterogeneous in terms of genre and domain, containing non-normative Russian usage, which increases the difficulty of the classification task.

3 The Feature Set

3.1 Dataset Preprocessing and the Context Windows

The window-dependent features described below (the semantic, the lexical, and the morphosyntactic ones) were computed (a) on the fixed context windows of the sizes 2, 3, 4, and 5; (b) on the unfixed-size window equivalent to the length of the full sentence; and (c) on the set of the syntactic arguments of the target verb (its direct dependencies and some of their secondary projections).

Only content non-stopwords were included into the semantic and the lexical windows; as for the morphosyntactic windows, they were comprised of all the grammemes found within a given window, including prepositions and punctuation marks.

The syntactic arguments of the target verbs and the morphological characteristics of lemmas were obtained with the online interface for the Russian MaltParser [7].

[2] The definitions throughout the paper are quoted from the Dictionary of the Russian Language [44].

3.2 Distributional Semantic Features

The Word-Embeddings Models. Our semantic features are based on word-embeddings models. We experiment with two pre-trained models presented in [19] that are freely available for download from the RusVectōrēs website [34]; both models were trained with the word2vec Continuous Skipgram algorithm.

- The WikiRNC model was trained with vector dimensionality 300 and window size 2 on the joint corpus of Russian Wikipedia and the Russian National Corpus, with the total of 600 m tokens;
- The Araneum model was trained on a much larger corpus, Araneum Russicum Maximum [5], of about 10bn tokens, with vector dimensionality 600 and the window size of 2.

The Semantic Similarity Measure. When we apply distributional semantics to context windows of different sizes, we proceed from the intuition that a metaphoric verb will be semantically deviant from its linear context window, affecting the mean semantic similarity between the words in the window in a negative way, whereas a literally used verb will belong to the same conceptual domain as its context words, making the contextual sub-space denser and adding to the mean similarity [14].

Application of distributional semantic models to the syntactic arguments of the verbs relies on the consideration that metaphor is a Selectional Preference violation [43], which is effectively captured as semantic deviance between the metaphoric verb and its main arguments [35]. The assumption is that a verb used in a literal sense will belong to the same conceptual domain as its immediate arguments, whereas metaphoric verb usage implies arguments belonging to a different conceptual domain.

The semantic similarity of tokens within the context is calculated as the following:

$$Sim_{win} = Mean\{\ Sim\left(w_i, w_j\right) | w_i, w_j \in Win\}, \tag{1}$$

$$SimV_{win} = Mean\{\ Sim\left(w_i; w_j\right) | w_i, w_j \in Win, w_i \neq verb, w_j \neq verb\}, \tag{2}$$

$$SimDiff_{win} = Sim_{win} - SimV_{win,} \tag{3}$$

where *Sim* is the semantic similarity in the distributional semantic space, and *Win* is the context window around the target verb: a linear window in the case of linear context, or the list of syntactic arguments in the case of the syntactic arguments context.

The Augmented Semantic Features. If a sentence in our corpus features a low-frequency word that is missing from the model, its measure of semantic similarity with its environment equals to zero. We moderated this effect by replacing the unavailable similarities by the mean of all the similarity measures in the current context window.

3.3 Lexical Co-occurrence Features

The use of lexical features for metaphor identification draws on the notion of lexico-semantic combinability [2], i.e. that different meanings of polysemous words impose restrictions on the semantics of their arguments, and subsequently, on their lexical classes. For example, the non-metaphoric meaning of *raspylyat* ‘to spray’ will often co-occur with lexemes from the class of liquids and powder-like substances (water, perfume, chemicals, and the like), while the metaphoric meaning ‘to scatter, to disperse smth thus decreasing its efficiency’ will typically co-occur with words denoting valuable resources (money, funds, effort, energy, troops, reserves, etc.).

To vectorize the unigrams of lemmas, we applied several measures of association: weirdness [1], the extension of Student’s t-test proposed in [24], log likelihood [6], and Kullback-Leibler Divergence [18]. The best results were produced by the ΔP metric [21] which is calculated according to the formula:

$$\frac{a}{a+\neg a}-\frac{b}{b+\neg b}, \tag{4}$$

where a is the number of occurrences of a lexeme in the metaphoric subcorpus, b is the number of occurrences of the lexeme in the non-metaphoric corpus, $a+\neg a$ is the size of the metaphoric subcorpus, and $b+\neg b$ is the size of the non-metaphoric subcorpus.

3.4 Morphosyntactic Co-occurrence Features

The rationale behind the use of morphosyntax in metaphor identification is grounded in the fact that different meanings of a polysemous verb may develop exclusive morphosyntactic constructions. For example, in the verb *otrubit* (whose non-metaphoric meaning is ‘to hack smth off’), the metaphoric meaning (‘to respond, to say smth in a brusque or abrupt manner’) develops an intransitive construction; this meaning is often used to introduce direct speech in the narration:

— Нет, — <отрубил> Керк. — Деньги должны быть выиграны сегодня. ‘No’, Kirk <cut off> (= responded abruptly), ‘the money must be won today’.

We explored three different configurations of morphological characteristics of nouns and verbs which vary in the fullness of representation:

1. verb only pos/noun only pos: indication of only the part of speech;
2. verb full: part of speech, aspect, tense, number, mood, gender, and person;
3. noun full: part of speech, gender, animacy, case, and number;
4. verb short: part of speech, aspect, tense, mood;
5. noun short: part of speech, animacy, case;

We tested five combinations of morphological configurations: verb only pos + noun only pos, verb full + noun full, verb short + noun full, verb full + noun short, and verb short + noun short. Prepositions and punctuation in all the configurations were represented by their lemmas; all the other parts of speech were always represented by their POS tags.

Besides, we experimented with unigrams, bigrams, and trigrams of morphosyntactic tags: bigrams and trigrams are expected to capture the linear order of grammemes in the context window, while unigrams show their distribution in sentences irrespective of the linear order. The association measure between grammemes on the one hand, and the non-metaphoric/metaphoric class on the other was calculated with the ΔP metric.

4 Experimental Setup

The metaphor identification task was formulated as sentence-level binary classification: the classifier was to identify which sentences belonged to the metaphoric and the non-metaphoric classes. We experimented with the datasets of individual verbs and with the combined dataset of all the 20 verbs.

We used the Support Vector Machine (SVM) classifier with linear kernel[3]; the experiments were run using 5-fold cross-validation.

We experimented with a total of 45 models, i.e. with different one-, two-, three-, four-, and five-feature combinations.

The results of the performance were estimated as the accuracy of classification.

5 Results

5.1 Features' Impact

Beside the features described in Sects. 3.2–3.4, we also tested the following features: (a) specific lexical markers of metaphoricity ('flag words', see [12, 36]); (b) quotation marks; and (c) sentence length. However, none of them proved efficient, either in isolation or in combination with the other features.

All the window-dependent features (semantic, lexical, and morphosyntactic) have proved to be quite sensitive to the size of the context window. Figure 1 demonstrates the correlation between the classification results (accuracy) on the lexical features and the size of the window for three the verbs which demonstrate a downward, an upward, and a flat dynamics.

Obviously, this behaviour is connected with the distances at which the lemmas with conspicuous association scores occur in relation to the target verb.

For example, *otrubit* 'to hack smth off' best performs on the set of the syntactic arguments; this is due to the high frequency of the metaphoric intransitive construction which serves to introduce direct speech (see Sect. 3.4). The verb in this construction has only one syntactic argument, the subject, which is typically a person's name. Proper names are low-frequency lemmas, and therefore they will have low association scores. Whereas the non-metaphoric meaning tends to co-occur with higher-frequency syntactic arguments on a much more regular basis (e.g. 'to cut off a chunk of wood/smb's head', etc.); these lemmas have high association scores. This contrast imparts high predictive power to the model based on the syntactic arguments of *otrubit*. Using linear windows, especially larger ones, introduces excessive noise into the model and disorients the classifier.

However, in the aggregate terms across the dataset, the large-size windows (full_sent and win5) by far outperform the other windows.

The accuracies of the non-augmented models and their augmented counterparts showed no significant difference.

---

[3] LinearSVC, as implemented in scikit-learn [32].

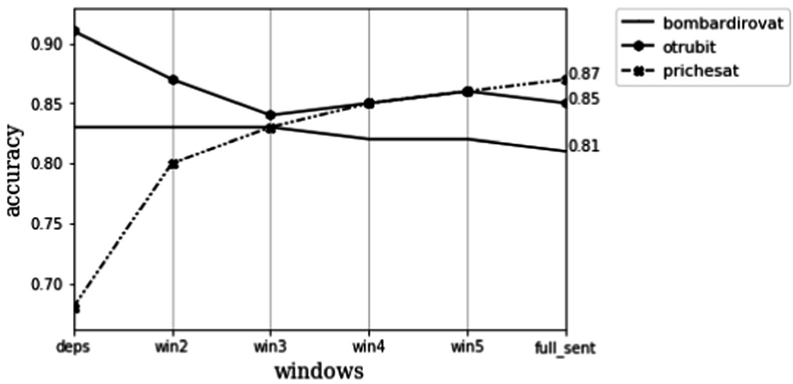

Fig. 1. Correlation between the accuracy of classification and the size/type of the context window (lexical co-occurrence features). '*Deps*' – the set of the verb's syntactic arguments; '*win2*' – '*win5*' – windows of the sizes 2–5; '*full_sent*' – window of the full sentence length.

The morphologically poor configuration of grammemes ('only pos') is demonstrably outperformed by the morphologically informed configurations (2–5, see the list in Sect. 3.4). Meanwhile, there is no pronounced leader among the morphologically informed configurations: they all perform at approximately the same level.

Besides, morphosyntactic unigrams consistently outperform trigrams, while being almost on a par with bigrams.

In sum, the efficient models provided by our features are the one-, two-, and three-feature combinations of the semantic, the lexical, and the morphosyntactic features.

5.2 Classification Results

We report the results for the models with the following options:

- the features are computed on the full sentence window;
- the distributional semantic feature ('sem') is the non-augmented version computed on the Araneum word-embeddings model;
- the morphosyntactic feature ('morph') is computed on unigrams of the configuration 'verb full + noun full';
- the lexical co-occurrence feature ('lex') is computed as described in Sect. 3.3.

An abridged version of the classification results is presented in Table 2. The full version of the table can be accessed online (See footnote 1).

The accuracies of the models across the verbs range within the following limits: 'sem': 0.52–0.81; 'lex': 0.77–0.94; 'morph': 0.67–0.82; 'sem+lex': 0.77–0.94; 'sem +morph': 0.7–0.85; 'lex+morph': 0.77–0.96; 'sem+lex+morph': 0.75–0.95.

The best accuracies on individual verbs range from the moderate 0.77 to the quite encouraging 0.96. The accuracy of the classifier on the combined dataset of the 20 verbs reached the mark of 0.83. This performance is on a competitive footing with the results reported by the other systems for metaphor identification in Russian: the F-scores of 0.76 in [39] and 0.84 in [38] which use the translation method and experiment

Table 2. Accuracy of classification (selected verbs)

dataset / model	sem	lex	morph	sem+lex	sem+morph	lex+morph	sem+lex+ morph
bombardirovat	0.75	0.81	0.75	0.83	0.77	0.82	0.85
napadat	0.59	0.77	0.76	0.77	0.74	0.77	0.75
vykraivat	0.81	0.94	0.82	0.93	0.85	0.96	0.95
combined dataset (20 verbs)	0.65	0.82	0.67	0.82	0.71	0.83	0.83

with much smaller datasets of pre-filtered SVO triples and adjective-noun tuples; and the accuracy of 0.68 in [31] which is run in a setting comparable to ours.

In five of the 20 verbs, the best result is achieved with the simple model 'lex'; adding further features does not lead to a gain in efficiency. The composite model of the semantic and the lexical features ('sem+lex') yields the best result only in two verbs. The majority of the top accuracies is achieved with the combination of the two features, the lexical and the morphosyntactic ones, ('lex+morph') – in 10 of the individual verbs, and on the joint dataset. In four individual verbs, the best results are obtained with the most complex model composed of the three features, the semantic, the lexical, and the morphosyntactic ones ('sem+lex+morph'). On the joint dataset, the last two models yield an identical result.

Interestingly, the 'sem', the 'morph', and the 'sem+morph' models consistently fall behind the other models across the datasets, as morphology alone cannot be expected to reliably predict the metaphoric or the non-metaphoric class. As for the comparatively low efficiency of the distributional semantic feature, it presumably can be accounted for by the fact that state-of-the-art distributional semantic models do not discriminate between different meanings of polysemous words; they generate a single vector which collapses all the senses of a word into a single value. The classification results will depend on the nature of the typical senses of the target verb and their co-occurrences in the training corpus (a fact also addressed in [31]).

To summarize, we can say that on a dataset composed of multiple target verbs, the two models are most likely to produce the high accuracy result: the two-feature combination 'lex+morph', and the three-feature combination 'sem+lex+morph'.

However, this observation may hold true only for verbs that are characterised by the semantic and the actant structure properties described in Sect. 2.1.

6 Conclusion

We have presented a manually annotated experimental dataset of metaphoric and non-metaphoric sentences featuring 20 target verbs. We also introduced the set of experimental features and presented their linguistic motivation. Next, we described the setup of the experiment for classifying the sentences into the metaphoric and the non-metaphoric classes. The results of the experiment suggest that the two composite models are likely to be scalable: the model combining the lexical and the morphosyntactic features, and the model based on the combination of the semantic, lexical, and morphosyntactic features. However, this generalization may hold true only for verbs of the same type as the target verbs in the experimental dataset (i.e. typical Activity or Accomplishment verbs with two actants).

Acknowledgements. The contribution to this study by Polina Panicheva is supported by RFBR grant № 16-06-00529.

References

1. Ahmad, K., Gillam, L., Tostevin, L.: Weirdness indexing for logical document extrapolation and retrieval (WILDER). In: Voorhees, E., Harman, D. (eds.) Proceedings of the 8th Text Retrieval Conference, TREC 8, Gaithersburg, MA, pp. 717–724 (2000)
2. Apresjan, Yu. D.: Izbrannyye trudy, t.1. Leksicheskaya semantika. Sinonimicheskiye sredstva yazyka/Selected works, vol. 1. Lexical semantics. The synonymic means of the language, 2nd edn. LRC Publishing House, Moscow (1995)
3. Beigman Klebanov, B., Leong, C.W., Gutierrez, E.D., Shutova, E., Flor, M.: Semantic classifications for detection of verb metaphors. In: Proceedings of the 54th Annual Meeting of ACL 2016, Berlin, Germany, vol. 2, pp. 101–106 (2016)
4. Beigman Klebanov, B., Leong, C.W., Heilman, M., Flor, M.: Different texts, same metaphors: unigrams and beyond. In: Proceedings of the Second Workshop on Metaphor in NLP, Baltimore, MD, pp. 11–17 (2014)
5. Benko, V., Zakharov, V.: Very large Russian corpora: new opportunities and new challenges. In: Selegey, V. (ed.) Proceedings of the Annual International Conference "Dialogue", Moscow, Russia, pp. 83–98 (2016)
6. Cressie, N., Read, T.R.C.: Multinomial goodness-of-fit tests. J. R. Stat. Soc. Ser. B (Methodol.) **46**(3), 440–464 (1984)
7. Droganova, K.A., Medyankin, N.S.: NLP pipeline for Russian: an easy-to-use web application for morphological and syntactic annotation. In: Proceedings of the Annual International Conference "Dialogue", Moscow, Russia (2016)
8. Dunn, J.: Evaluating the premises and results of four metaphor identification systems. In: Proceedings of CICLing 2013, Samos, Greece, pp. 471–486 (2013a)
9. Dunn, J.: What metaphor identification systems can tell us about metaphor-in-language. In: Proceedings of the First Workshop on Metaphor in NLP, Atlanta, GA, pp. 1–10 (2013b)
10. Gandy, L., et al.: Automatic identification of conceptual metaphors with limited knowledge. In: Proceedings of the 27th AAAI Conference on Artificial Intelligence, pp. 328–334. AAAI Press, Bellevue, WA (2013)

11. Gedigian, M., Bryant, J., Narayanan, S., Ciric, B.: Catching metaphors. In: Proceedings of the Third Workshop on Scalable Natural Language Understanding, ScaNaLU 2006, New York City, pp. 41–48 (2006)
12. Goatly, A.: The Language of Metaphors. Routledge, Abingdon (2011)
13. Heintz, I., et al.: Automatic extraction of linguistic metaphors with LDA topic modeling. In: Proceedings of the First Workshop on Metaphor in NLP, Atlanta, GA, pp. 58–66 (2013)
14. Herbelot, A., Kochmar, E.: 'Calling on the classical phone': a distributional model of adjective-noun errors in learners' English. In: Proceedings of COLING 2016, Osaka, Japan, pp. 976–986 (2016)
15. Hovy, D., et al.: Identifying metaphorical word use with tree kernels. In: Proceedings of the First Workshop on Metaphor in NLP, Atlanta, GA, pp. 52–57 (2013)
16. Kilgarriff, A., et al.: The Sketch Engine: ten years on. Lexicography **1**, 7–36 (2014)
17. Krishnakumaran, S., Zhu, X.: Hunting elusive metaphors using lexical resources. In: Proceedings of the Workshop on Computational Approaches to Figurative Language, NAACL-HLT 2007, Rochester, NY, pp. 13–20 (2007)
18. Kullback, S., Leibler, R.A.: On Information and Sufficiency. Annals of Mathematical Statistics, vol. 22(1), pp. 79–86. Institute of Mathematical Statistics, Ann Arbor (1951)
19. Kutuzov, A., Kuzmenko, E.: WebVectors: a toolkit for building web interfaces for vector semantic models. In: Ignatov, D.I., et al. (eds.) AIST 2016. CCIS, vol. 661, pp. 155–161. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52920-2_15
20. Lakoff, G., Johnson, M.: Metaphors We Live By. The University of Chicago Press, Chicago (1980)
21. Levshina, N.: How to Do Linguistics with R: Data Exploration and Statistical Analysis. John Benjamins Publishing Company, Amsterdam (2015)
22. Li, L., Sporleder, C.: Classifier combination for contextual idiom detection without labelled data. In: Proceedings of EMNLP 2009, Singapore, pp. 315–323 (2009)
23. Li, L., Sporleder, C.: Using gaussian mixture models to detect figurative language in context. In: Proceedings of NAACL HLT 2010, Los Angeles, CA, pp. 297–300 (2010)
24. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge (1999)
25. Mason, Z.J.: CorMet: a computational, corpus-based conventional metaphor extraction system. Comput. Linguist. **30**(1), 23–44 (2004)
26. Mehlig, H.R.: Semantika predlozheniya i semantika vida v russkom yazyke/The semantics of the sentence and the semantics of the aspect in the Russian language. In: Bulygina, T.V., Kibrik, A.E. (eds.) Novoye v zarubezhnoy lingvistike, vyp. 15. Sovremennaya zarubezhnaya lingvistika/The state-of-the-art in international linguistics, issue 15. Contemporary international linguistics, pp. 227–249. Progress, Moscow (1985)
27. Mohler, M., Bracewell, D., Tomlinson, M., Hinote, D.: Semantic signatures for example-based linguistic metaphor detection. In: Proceedings of the First Workshop on Metaphor in NLP, Atlanta, GA, pp. 27–35 (2013)
28. Mohler, M., Rink, B., Bracewell, D.B., Tomlinson, M.T.: A novel distributional approach to multilingual conceptual metaphor recognition. In: Proceedings of COLING 2014: Technical Papers, Dublin, Ireland, pp. 1752–1763 (2014)
29. Neuman, Y., et al.: Metaphor identification in large texts corpora. PLoS ONE **8**(4), e62343 (2013)
30. Ovchinnikova, E., Israel, R., Wertheim, S., Zaytsev, V., Montazeri, N., Hobbs, J.: Abductive inference for interpretation of metaphors. In: Proceedings of the 2nd Workshop on Metaphor in NLP, NAACL-HLT 2014, Denver, CO, pp. 33–41 (2014)

31. Panicheva, P., Badryzlova, Yu.: Distributional semantic features in Russian verbal metaphor identification. In: Selegey, V. (ed.) Proceedings of the Annual International Conference “Dialogue”, vol. 1, pp. 179–190. Moscow, Russia (2017)
32. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
33. Peters, W., Peters, I.: Lexicalised systematic polysemy in WordNet. In: Proceedings of LREC 2000, Athens, Greece (2000)
34. RusVectōrēs: Word Embeddings for Russian Online. http://rusvectores.org/ru/models/. Accessed 01 July 2018
35. Shutova, E., Kiela, D., Maillard, J.: Black holes and white rabbits: metaphor identification with visual features. In: Proceedings of NAACL HLT 2016, pp. 160–170 (2016)
36. Steen, G.J., Dorst, L., Herrmann, B., Kaal, A., Krennmayr, T., Pasma, T.: A Method for Linguistic Metaphor Identification: from MIP to MIPVU. John Benjamins Publishing, Amsterdam (2010)
37. Strzalkowski, T., et al.: Robust extraction of metaphor from novel data. In: Proceedings of the First Workshop on Metaphor in NLP, Atlanta, GA, pp. 67–76 (2013)
38. 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 ACL, vol. 1, Baltimore, MD, pp. 248–258 (2014)
39. Tsvetkov, Y., Mukomel, E., Gershman, A.: Cross-lingual metaphor detection using common semantic features. In: Proceedings of the First Workshop on Metaphor in NLP, Atlanta, GA, pp. 45–51 (2013)
40. Turney, P.D., Neuman, Y., Assaf, D., Cohen, Y.: Literal and metaphorical sense identification through concrete and abstract context. In: Proceedings of EMNLP 2011, Edinburgh, UK, pp. 680–690 (2011)
41. Veale, T., Shutova, E., Beigman Klebanov, B.: Metaphor: a computational perspective. In: Synthesis Lectures on Human Language Technologies, vol. 9, no. 1, pp. 1–160 (2016)
42. Vendler, Z.: Verbs and times. Philos. Rev. **66**, 143–160 (1957)
43. Wilks, Y.: Making preferences more active. Artif. Intell. **11**(3), 197–223 (1978)
44. Yevgenyeva, A.P. (ed.): Slovar russkogo yazyka v 4 tomakh/The dictionary of the Russian language in 4 volumes, 2nd edn. The Institute for the Russian Language, Moscow (1981–1984)