

Detection of Semantic Changes in Russian Nouns with Distributional Models and Grammatical Features

Anastasiia Ryzhova
Federal Research Center
"Computer Science and Control"
of the Russian Academy
of Sciences,
Lomonosov Moscow
State University
Moscow, Russia
ryzhova@tesyan.ru

Daria Ryzhova
HSE University

Moscow, Russia
dryzhova@hse.ru

Ilya Sochenkov
HSE University

Moscow, Russia
isochenkov@hse.ru

Abstract

The paper presents the models detecting the degree of semantic change in Russian nouns developed by the team *aryzhova* within the RuShiftEval competition of the Dialogue 2021 conference. We base our algorithms mostly on unsupervised distributional models and additionally test a model that uses vectors representing morphological preferences of the words in question. The best results are obtained by the model built on the ELMo architecture with a small window, while the quality of performance of the "grammatical" model is comparable to that of the models based on much more sophisticated algorithms.

Keywords: semantic change, COMPARE metric, distributional models, word2vec, ELMo, RuBERT, grammatical profile

DOI: 10.28995/2075-7182-2021-20-597-606

Оценка степени семантических изменений у русских существительных с помощью дистрибутивных моделей и грамматических профилей

Анастасия Александровна Рыжова
Федеральный исследовательский центр «Информатика и управление» Российской академии наук, МГУ имени М.В. Ломоносова
Москва, Россия
ryzhova@tesyan.ru

Дарья Александровна Рыжова
Национальный исследовательский университет «Высшая школа экономики»
Москва, Россия
dryzhova@hse.ru

Илья Владимирович Соченков
Национальный исследовательский университет «Высшая школа экономики»
Москва, Россия
isochenkov@hse.ru

Аннотация

В статье представлены методы оценки степени семантических изменений русскоязычных существительных, разработанные в рамках соревнования RuShiftEval (Диалог-21) командой *aryzhova*. В качестве основы используются дистрибутивные модели в различных архитектурах (word2vec, ELMo, RuBERT), а также грамматические профили – вектора частотностей морфологических форм, в которых встречаются в корпусах разных временных периодов анализируемые существительные. Лучшие результаты показывает модель на базе архитектуры ELMo, учитывающей также ближайший контекст (окно = 1), а модель на основе одних только грамматических профилей дает результаты, сопоставимые с показателями значительно более сложных алгоритмов.

Ключевые слова: семантический сдвиг, метрика COMPARE, дистрибутивные модели, word2vec, ELMo, RuBERT, грамматический профиль

1 Introduction

The paper presents a series of experiments conducted by the team *aryzhova* within the Dialogue’21 RuShiftEval competition aimed at automatic detection of a degree of semantic change in Russian nouns throughout three time periods: a pre-Soviet (1700-1916), a Soviet (1918-1990), and a post-Soviet (1991-2016) ones [20].

Semantic change is a shift in the meaning of a lexeme occurring due to socio-cultural or purely linguistic reasons. Such changes can be substantial and clearly seen. A common example is a new technical device labeled with an already existing lexical item, cf. a computer *mouse*, or Russian *mama* ‘mother’ acquiring the meaning ‘motherboard’. At the same time, some changes affect just a particular aspect of a word’s meaning, such as its general connotations and associations, cf. *Kosovo* which is strongly associated with war thematics after the conflicts in 1998-1999 [21]. Between these extremities, there are plenty of intermediate cases.

The RuShiftEval shared task consisted in ranging a set of Russian nouns according to the degree of semantic change which they underwent in a given time span. We test a set of distributional models (word2vec [22], ELMo [25], RuBERT [16]) with different parameter settings in application to this task.

The aim of our study is twofold. On the one hand, we intend to test the applicability of the most prominent semantic change detection methods that have proven their efficiency in languages other than Russian (mostly in English). Since Russian, in contrast to English, has a rich morphology, we make several attempts to take it into consideration. On the other hand, we expect the results of the experiments to shed new light on the linguistic nature of semantic change. Thus some of our experiments test certain linguistic properties of the given lexical items.

2 Related work

Semantic change is a linguistic phenomenon intriguing scholars from a long time ago (see, e.g. [3]). However, as a specific domain of studies, it shaped relatively recently. The main achievements in the domain are still mostly limited to the detailed descriptions of individual words’ trajectories of semantic change (cf. [1]; [14], among others) or the research of individual mechanisms of semantic shift, such as metaphor, metonymy, and others ([15]; [9], to mention just a few).

The emergence of large-scale corpora gave rise to computational studies of semantic change ([17]; [6]) and allowed to formulate and test some general laws underlying these semantic processes. Among these general laws are a correlation between a word’s frequency, its level of polysemy, and its aptitude to meaning change [11], or dependency of the degree of semantic change on the level of the word’s prototypicality in the previous time period [7]. However, the lion’s share of the studies is still based solely on the English data.

In Russian linguistic tradition, there is a fundamental study by V. V. Vinogradov [31] describing an impressive number of words and expressions through the prism of their diachronic change, followed by a recent volume [5] representing an in-depth corpus analysis of twenty words across two centuries. Due to the lack of diachronic corpus data in open access, computational analysis of semantic change in Russian was limited until very recently. Now, the release of the diachronic subcorpora of the RNC opens up new perspectives in the field.

The most prominent methodologies of semantic change detection in English (and some other languages) make use of various distributional models [17]. We use these algorithms as a basis of our experiments as well.

3 Dataset

The organizers of the competition provided three datasets: train, development, and test.

The train set consists of two parts, called RuSemShift1 and RuSemShift2, the first semantic change datasets in Russian [27]. The RuSemShift1 covers pre-Soviet and Soviet times and includes 48 words; the RuSemShift2 covers Soviet and post-Soviet times with 51 words. The datasets contain both nouns and adjectives. However, we exclude the adjectives from consideration since the development and test sets

Dataset	Number of tokens
pre-Soviet	73542513
Soviet	95043479
post-Soviet	83269542

Table 1: The number of tokens in RNC texts.

contain only nouns. The resulting datasets include 44 and 43 words in the RuSemShift1 and RuSemShift2, respectively.

We used recently released diachronic sub-corpora of the Russian National Corpus for our experiments, which correspond to the three time periods. Table 1 presents the volumes of these corpora in tokens.

The semantic change value is measured by COMPARE metric [29]. In brief, for each word, the annotators get random pairs of sentences. The sentences belong to two different periods. The annotators evaluate them, placing scores from 1 to 4, where the lowest score corresponds to the strongest change. The resulting score for each pair of sentences is the average of the scores of the annotators. The score for each word is the average of the scores for each pair of sentences.

The development dataset includes 12 words, the test dataset – 99 words. For each word the participants of the competition have to predict semantic change score (COMPARE) in three pairs of time periods: pre-Soviet - Soviet (RuSemShift1), Soviet - post-Soviet (RuSemShift2), pre-Soviet - post-Soviet (RuSemShift3). The words are ranged according to these scores, and the results are evaluated with the Spearman correlation between the produced ranking and the ranking obtained from human annotation.

4 Experiments

Our main experiments are based on the model architectures suggested in [17]. We use static word embeddings, word2vec [22], and contextualized word embeddings, ELMo [25]. The word2vec model considers only corpus statistics, but it can capture some semantic properties of words, assigning to each word exactly one vector representation. The ELMo model has the BiLSTM architecture and is believed to catch deeper semantic properties. The embedding of each word depends on the given context, so for each word mention the model provides a separate embedding vector. In our research, we consider models trained both on lemmas and on tokens. In addition, we conduct a simplistic experiment testing whether a change in a morphological profile correlates with a change in meaning.

Below we describe each experimental setup in more detail.

4.1 Experiment 1: Word2vec

In the first experiment series, we use the word2vec models provided by the organizers of the competition. These models are trained separately on the Russian diachronic corpora using CBOW algorithm, context window size equals 5, vector size is 300.

As a baseline, we range the words by the cosine similarity of their word2vec representations in the models corresponding to the different time periods. The models were trained on lemmas and aligned with Procrustes alignment (cf. [11]).

On the word2vec models trained on tokens we run three different experiments. We compute final scores in the same way as in the baseline, but we test three types of vector representations:

1. Vector of the word is the average vector of all its word forms.
2. Vector of the word is the vector of its most frequent word form.
3. Vector of the word is the vector of its word form which displays the highest rate of semantic change in the given period. We computed the cosine similarity of word embeddings for each word form, and then took the one with the lowest final score.

4.2 Experiment 2: ELMo

The following experiment series is based on ELMo contextualized embeddings.

For each word we find all the sentences where these words occur and construct 100 random pairs. The first sentence in a pair belongs to one time period, the second to the other period. For example, in the RuSemShift1 task, the first sentence is picked from the pre-Soviet time corpus, the second – from the Soviet one. For each sentence we compute the ELMo embedding of the word for which we want to measure the semantic shift value. We also build additional models where each sentence is presented by the average vector of the target word embedding and the embeddings of the words from its closest context (one, two, or three items away from the target word), which proved efficient in similar experiments from [18]. The final score is the average cosine similarity for all pairs. To achieve a certain level of robustness, we take the average value of the cosine similarity from 5 experiments on different 100 random pairs in each iteration.

Another parameter that we vary in this experiment series is sentence preprocessing: the target and the context words are either treated as tokens or substituted by their lemmas. For the experiment on tokens, we use the tokens ELMo model *ruwikiruscorpora_tokens_elmo_1024_2019*, trained on RNC and Wikipedia, from RusVectores [19]. For the experiment on lemmas, we lemmatize texts with the *morpho_ru_syntagrus_pymorphy* model from DeepPavlov library, which also allows excluding personal names from consideration. It is a neural morphological tagger; the algorithm is described in [13]. The ELMo model *ruwikiruscorpora_lemmas_elmo_1024_2019* is also taken from RusVectores.

4.3 Experiment 3: RuBERT model

This experiment repeats the previous algorithm with another model, RuBERT, presented in [16]. It is a standard BERT model with transformer architecture, trained on the Russian Wikipedia and news, based on the multilingual BERT model. In this case we take into account only the target word embedding.

4.4 Experiment 4: Grammatical features

Previous theoretical studies of various semantic phenomena show that a meaning change can affect not only the distributional properties of a word but its grammatical profile as well (see, for example, [4]). Trying to test this hypothesis on large-scale data, we represent each target noun with a vector of its grammatical preferences. These grammatical vectors consist of 12 dimensions that correspond to different morphological forms appropriate for Russian nouns, i.e., the combinations of six cases and two grammatical numbers. The values of these dimensions are computed as raw counts of tokens of the target word in the given morphological form within each time period. The final scores, which are expected to signal the level of semantic change, are computed as in the previous experiment series – as the cosine similarity of two vectors representing the same word in different time spans.

4.5 Experiment 5: Regression

In the final experiment, we try to benefit from distributional and grammatical properties of words taken together. To achieve this, we train a linear regression with regularization given by the l2-norm (Ridge Regression). The regularization strength alpha is selected with the gridsearch on 5 fold cross-validation and equals to 0.1. We use the cosine similarities from the experiment in Section 4.2 with an ELMo model (tokens + context of window size 1), and the cosine similarities of the corresponding grammatical vectors as features.

5 Results

We conducted all our experiments on the train RuSemShift1 dataset, and then evaluated the best methods on other datasets. Tables 2-6 present the results we obtained. In Tables 5 and 6, *Spearman correlation 1*, *Spearman correlation 2*, and *Spearman correlation 3* state for correlation coefficients corresponding to the three time period pairs: pre-Soviet and Soviet, Soviet and post-Soviet, pre-Soviet and post-Soviet, respectively; an asterisk in the same tables indicates the results where the p-value is lower than 0.05.

Model	Spearman correlation
word2vec similarity	0.485
ELMo, layers='average'	0.508
ELMo, layers='top'	0.526
ELMo with context, 'average', window=2	0.579
ELMo with context, 'top', window=2	0.555

Table 2: Results on the training set RuSemShift1 (44 nouns), lemmatized texts

Model	Spearman correlation
word2vec, average of all word forms	0.282
word2vec, the most frequent word form	0.361
word2vec, the most changed word form	0.28
ELMo, layers = 'average'	0.54
ELMo with context, 'average', window=3	0.593
ELMo with context, 'average', window=2	0.593
ELMo with context, 'average', window=1	0.621
RuBERT	0.411
grammatical vectors	0.30
linear regression	0.602

Table 3: Results on the training set RuSemShift1 (44 nouns), tokens

Tables 2-6 show that the results differ depending on the dataset. However, it is clearly seen that the ELMo model with a small window size outperforms both word2vec and ELMo with a bigger window size models. RuBERT performs worse than ELMo, but it should be taken into account that the ELMo models were trained on the RNC collection, while we do not fine-tune RuBERT on the same corpora.

Interestingly, pretty simple grammatical vectors get rather high scores, even outperforming word2vec and RuBERT models on some datasets. It means that change in meaning is indeed correlated with grammatical re-profiling. Figure 1a, representing the distribution of morphological forms of the noun *svalka* in the pre-Soviet and the Soviet subcorpora, gives an illustration of a clear grammatical shift which can be easily explained from the semantic point of view. The most frequent morphological form of this word in the pre-Soviet period is the nominative singular, while in the Soviet period, the accusative singular form becomes almost as frequent as the nominative singular one. The prevalent meaning of this word in the pre-Soviet subcorpus was that of a fight [24]. Since the word denoted an event, it was frequently used in existential contexts, declaring that a fight was taking place (see Example 1) – hence the preference for the nominative case. In contrary, in the Soviet period this meaning is rarely attested, giving way to the meaning 'dumping ground'. This new semantics triggers the usages of the word *svalka* in the accusative case, because, denoting a specific place, it often plays the role of the goal of a motion (Example 2) typically marked with the accusative case in Russian.

- (1) *Totčas že na zemle zakipela svalka.NomSg, i desyatki tel smešalis' v odnu obš'uyu kričaš'uyu massu.*

'A **scuffle** ensued, with dozens of women in a bawling, struggling mass on the ground.'
[Aleksandr Kuprin. Olesya (Stepan Apresyan, 1982)]

- (2) *Govoryat šefu: stanok slomalsya. On verit, volokut stanok na svalku.AccSg.*

'They would tell the boss that a lathe was broken. He would believe them and they would drag the lathe out on to the *rubbish dump*.' [Anatoly Kuznetsov. Babi Yar (David Floyd, 1970)]

Model	Spearman correlation
word2vec on lemmas	0.545
ELMo tokens with context, layers='average', window =1	0.715
RuBERT	0.211
grammatical vectors	0.465
linear regression	0.733

Table 4: Results on the training set RuSemShift2 (43 nouns)

Model	Spearman correlation 1	Spearman correlation 2	Spearman correlation 3
word2vec on lemmas	0.49	0.538	0.622*
ELMo lemmas, 'average'	0.559	0.343	0.664*
ELMo tokens + context, 'average', window =1	0.636*	0.769*	0.818*
RuBERT	0.322	0.531*	0.517*
grammatical vectors	0.392	0.259	0.252
linear regression	0.741*	0.727*	0.832*

Table 5: Results on the development dataset (12 nouns)

Model	Spearman correlation 1	Spearman correlation 2	Spearman correlation 3
word2vec on lemmas	0.141	0.246*	0.330*
ELMo lemmas, 'average'	0.469*	0.450*	0.453*
ELMo tokens + context, 'average', window =1	0.430*	0.451*	0.469*
RuBERT	0.380*	0.429*	0.448*
grammatical vectors	0.157	0.199*	0.343
linear regression	0.480*	0.487*	0.560*

Table 6: Results on the test dataset (99 nouns)

It is also clear that change in morphological preferences is more crucial in some cases, while in others they seem insignificant. For example, the noun *element* 'element' exhibits substantial semantic differences between the pre-Soviet and the Soviet periods, according to the annotators (its COMPARE equals 1.91), nevertheless it remains grammatically stable.

The opposite situation, where grammatical re-profiling is present, while almost no semantic change is detected (i.e., COMPARE value is higher than 3) is also attested. This is usually the case for rare words, such as *roždestvo* 'Christmas' or *agenstvo* 'agency'. However, it seems that, for some words, grammatical changes reveal some interesting tendencies remained unnoticed by the annotators. For example, the word *pravitel'*, which does not change in meaning from the pre-Soviet to the Soviet times, according to human annotation (COMPARE equals 3.38), shows a clear shift towards plural forms with almost no change in the case forms ratio. This phenomenon could reflect a cultural (political) change in the corresponding linguistic society, i.e. the shift from monarchy to socialism. However, such effects require further investigation.

As for the regression model, it only slightly outperforms other models, being in general compatible with

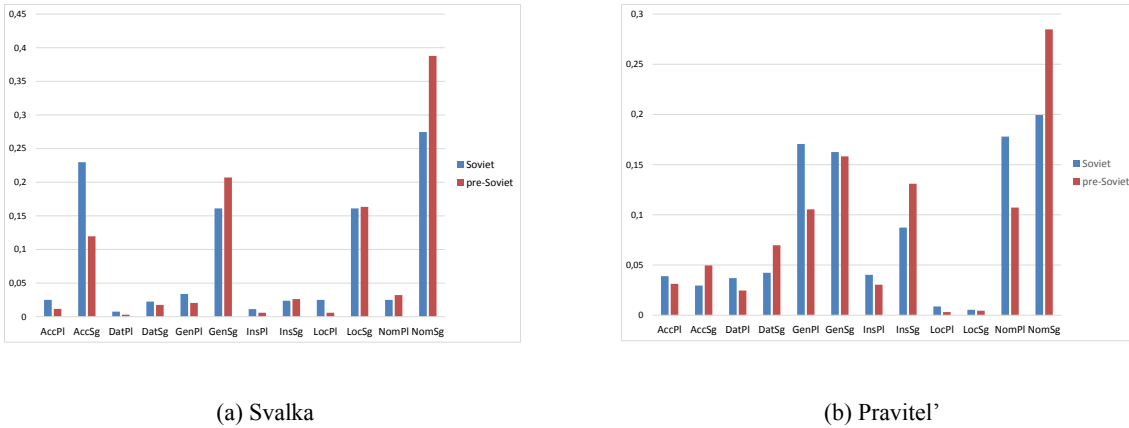


Figure 1: Changes in grammatical profiles of the words *svalka* and *pravitel'* between the pre-Soviet and the Soviet periods.

the best ELMo model. However, it seems to add some robustness to the algorithm and shows substantial improvement on the datasets with words that have undergone more or less significant grammatical re-profiling.

6 Discussion

Most evidently, the lack of textual data leads to a worse model performance, and this factor is expectedly more crucial for models based on the word2vec CBOW architecture than for those based on the ELMo algorithm. The pre-Soviet subcorpus is the most problematic in this respect: while covering the longest time span, it is the most modest in size, and it is less representative in terms of the text genres. In addition to it, this subcorpus is quite heterogeneous and seems to split into two parts: the texts written app. before the 1830s versus the texts written app. after 1830. This borderline is mentioned in a range of case studies of semantic changes occurred in individual lexical items (cf. [30] on *znatnyj* ‘famous’, [23] on *mama* ‘mother’, [26] on *slavnyj* ‘famous’, among others). It is grounded on observations that it is only after the 1830s that the traces of colloquial speech - the register which is mostly prone to semantic changes - start to penetrate into written texts reflected in the corpus. Thus, many words are used rather differently in these parts of the pre-Soviet subcorpus. This piece of information is lost when a word representation is built on this subcorpus as a whole. It could be better to divide this time period (and, consequently, the subcorpus) into two parts, or even totally exclude the texts written before the 1830s from consideration. However, the amount of textual data representing this period, being already insufficient, would decrease even more.

Another problematic area is the evaluation metrics. The ‘golden standard’ dataset that was used in the competition is compiled from human judgements on the level of semantic differences within pairs of usages of one and the same lexical item randomly picked from the corresponding text corpora (see [29] and Section 2). This benchmark has many advantages. First, annotators deal with a word in context, which allows them to estimate semantic similarity more accurately than it is done in traditional datasets on semantic similarity, such as WordSim353 [8], where words are given in isolation. Second, the annotation procedure is not very difficult and allows for crowdsourcing. Third, this approach to construction of an evaluation dataset is fully data-driven, it is not based on the previous knowledge acquired from dictionaries or other resources. Finally, because of the random sampling of the context pairs, the dataset roughly represents the respective frequencies of different word senses in the corpus.

At the same time, this benchmark has some drawbacks, the most important of which, to our mind, concerns the vague nature of semantic change. Semantic change is an umbrella term for very differ-

ent processes, including metaphoric and metonymic shifts, grammaticalization and pragmaticalization, specification and generalization of a word meaning, cf. [2], [9]. Heterogeneity of these processes complicates the task of annotating and ranking words according to the level of semantic change they have undergone, since word meanings change in very different ways and aspects. The complexity of the task for humans is clearly seen from the level of inter-annotator agreement, which sometimes does not exceed 0.2.

It might be fruitful to develop a classification of linguistic phenomena related to semantic change and to annotate evaluation datasets according to it, assigning different weights to different classes (e.g., 1 for metonymy, 2 for metaphor, 3 for grammaticalization, etc.). For every lexical item within a pair of time periods, such an annotation would include the list of the attested types of semantic changes together with the total scores counted as the sum of these types weights.

It should be admitted that construction of such a dataset would be pretty costly: it would require a certain level of annotators' linguistic proficiency and a more rigorous analysis of various data sources. However, it could be useful for theoretical studies of semantic change, since it would allow for testing different methodologies for different kinds of processes (cf. different metrics for culturally versus linguistically driven shifts in [10]). The results of such experiments could help to reveal new regularities appropriate for various linguistic phenomena.

For example, it is already well known that a change in meaning usually correlates with a change of the usage context. This assumption underlies the so-called distributional hypothesis [12, 28], which, in its turn, gives rise to distributional semantics, i.e. to the most prominent methodology used to complete the task of semantic change estimation. However, our experiments show that the correlation between the level of semantic change and the level of change of the word's morphological profile is sometimes also rather high. This effect is expected for grammaticalization and pragmaticalization processes, but the examples of such changes were not so numerous in the datasets. It seems that some grammatical effects are appropriate for semantic shifts of other types as well – this topic deserves further investigation.

7 Conclusion

The task of an automated semantic shift detection in Russian is a promising field for future experiments, interesting for both computational and theoretical linguistics. The methods that we had implemented did not receive the highest scores in the RuShiftEval competition. Our model based on the ELMo algorithm with the smallest window took the 6th place among 14 participating systems, while the regression model was not submitted. There is definitely much room for further improvement: additional corpus data, fine-tuning of the RuBERT model, or usage of the train dataset for a supervised model generation could result in better performance.

One of the most interesting theoretical outcomes that we got is a rather strong correlation between grammatical re-profiling and semantic change. We find it an interesting topic for further research which could shed additional light on the nature of semantic shifting.

We would also like to highlight that the competition was devoted to nouns, and it would be an interesting challenge to look for the best ways of semantic change detection in adjectives and verbs. We hope that the next competition within the Dialog conference will include this task.

Acknowledgements

We deeply thank Lidia Pivovarova and Andrey Kutuzov for the organization of the competition and for their constant support of its participants, as well as the anonymous Reviewers for their insightful comments which helped us to improve this text in many respects. We are also grateful to Ekaterina Rakhilina for the idea that grammar matters. Anastasiia Ryzhova and Ilya Sochenkov are supported by RFBR, research project no. 18-29-03187, and partially by the Ministry of Science and Higher Education of the Russian Federation according to the agreement between the Lomonosov Moscow State University and the Foundation of project support of the National Technology Initiative No 7/1251/2019 dated 15.08.2019 within the Research Program “Center of Big Data Storage and Analysis” of the National Technology Initiative Competence Center (project “Text mining tools for big data”). Daria Ryzhova is supported by RFBR, research project no. 20-012-00240.

References

- [1] Blank Andreas. Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. — De Gruyter Mouton, 2013.
- [2] Bloomfield L. Language. — London : Allen & Unwin, 1933.
- [3] Bréal Michel. Essai de sémantique (2nd ed.). — Paris, France : Hachette, 1899.
- [4] Divjak Dagmar, Gries Stefan Th. Ways of trying in Russian: Clustering behavioral profiles // *Corpus Linguistics and Linguistic Theory*. — 2006. — Vol. 2, no. 1. — P. 23–60.
- [5] Dobrushina Nina R. Daniel Mikhail A. Danova Margarita K. Opachanova Anastasiia S. Pechurina Varvara S. Skorinkin Daniil A. ...Sheshenina Aleksandra V. Two centuries in twenty words [Dva veka v dvadcati slovah]. — the National Research University Higher School of Economics, 2016.
- [6] Dubossarsky Haim. Semantic change at large: A computational approach for semantic change research : Ph. D. thesis / Haim Dubossarsky ; Ph. D. thesis, Hebrew University of Jerusalem, Edmond and Lily Safra Center for Brain Sciences. — 2018.
- [7] Dubossarsky Haim, Weinshall Daphna, Grossman Eitan. Outta control: Laws of semantic change and inherent biases in word representation models // *Proceedings of the 2017 conference on empirical methods in natural language processing*. — 2017. — P. 1136–1145.
- [8] Finkelstein Lev G., Matias Evgeniy et al. Placing Search in Context: The Concept Revisited // *ACM Transactions on Information Systems*. — 2002. — Vol. 20, no. 1. — P. 116–131.
- [9] Geeraerts Dirk. Diachronic prototype semantics: A contribution to historical lexicology. — Oxford University Press, 1997.
- [10] Hamilton William L, Leskovec Jure, Jurafsky Dan. Cultural shift or linguistic drift? comparing two computational measures of semantic change // *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing / NIH Public Access*. — Vol. 2016. — 2016. — P. 2116.
- [11] Hamilton William L, Leskovec Jure, Jurafsky Dan. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change // *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. — 2016. — P. 1489–1501.
- [12] Harris Zellig S. Co-occurrence and transformation in linguistic structure // *Language*. — 1957. — Vol. 33, no. 3. — P. 283–340.
- [13] Heigold Georg, Neumann Guenter, van Genabith Josef. An extensive empirical evaluation of character-based morphological tagging for 14 languages // *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. — 2017. — P. 505–513.
- [14] Hopper Paul J, Traugott Elizabeth Closs. Grammaticalization. — Cambridge, UK : Cambridge University Press, 2003.
- [15] Kövecses Zoltán. Metaphor: A practical introduction. — New York : Oxford University Press, 2010.
- [16] Kuratov Yuri, Arkhipov Mikhail. Adaptation of deep bidirectional multilingual transformers for Russian language // *Komp'yuternaja Lingvistika i Intellektual'nye Tehnologii*. — 2019. — P. 333–339.
- [17] Kutuzov Andrey. Distributional word embeddings in modeling diachronic semantic change : Ph. D. thesis / Andrey Kutuzov ; Ph. D. thesis, University of Oslo. — 2020.
- [18] Kutuzov Andrey, Giulianelli Mario. UiO-UvA at SemEval-2020 Task 1: Contextualised Embeddings for Lexical Semantic Change Detection // *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. — 2020. — P. 126–134.

- [19] Kutuzov Andrey, Kuzmenko Elizaveta. WebVectors: a toolkit for building web interfaces for vector semantic models // International Conference on Analysis of Images, Social Networks and Texts / Springer. — 2016. — P. 155–161.
- [20] Kutuzov Andrey, Pivovarova Lidia. RuShiftEval: a shared task on semantic shift detection for Russian // Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference. — 2021.
- [21] Kutuzov Andrei, Velldal Erik, Øvrelid Lilja. Tracing armed conflicts with diachronic word embedding models. — 2017. — 01. — P. 31–36.
- [22] Mikolov Tomas, Sutskever Ilya et al. Distributed Representations of Words and Phrases and their Compositionality // Advances in Neural Information Processing Systems. — 2013. — 10. — Vol. 26.
- [23] Opachanova Anastasiia S., Dobrushina Nina R. Mother [Mama] // Two centuries in twenty words [Dva veka v dvadcati slovah]. — the National Research University Higher School of Economics, 2016. — P. 72–93.
- [24] Pechurina Varvara S., Dobrushina Nina R. Dumping ground [Svalka] // Two centuries in twenty words [Dva veka v dvadcati slovah]. — the National Research University Higher School of Economics, 2016. — P. 317–338.
- [25] Peters Matthew, Neumann Mark et al. Deep Contextualized Word Representations // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). — 2018. — P. 2227–2237.
- [26] Rakhilina Ekaterina V., A. Ryzhova Daria. Microhistory of semantic shifts: the case of the Russian adjective slavnyj [Slavnii korabl' - omulevaia bochka. K mikroistorii semanticheskikh perehodov] // Proceedings of the Vinogradov Institute of the Russian Language/Trudi instituta ruskogo iazyka im. V.V. Vinogradova. — 2019. — Vol. 20. — P. 241–256.
- [27] Rodina Julia, Kutuzov Andrey. RuSemShift: a dataset of historical lexical semantic change in Russian // Proceedings of the 28th International Conference on Computational Linguistics. — 2020. — P. 1037–1047.
- [28] Sahlgren Magnus. The distributional hypothesis // Italian Journal of Disability Studies. — 2008. — Vol. 20. — P. 33–53.
- [29] Schlechtweg Dominik, im Walde Sabine Schulte, Eckmann Stefanie. Diachronic Usage Relatedness (DURel): A Framework for the Annotation of Lexical Semantic Change // Proceedings of NAACL-HLT. — 2018. — P. 169–174.
- [30] Skorinkin Daniil. Famous [Znatnyj] // Two centuries in twenty words [Dva veka v dvadcati slovah]. — the National Research University Higher School of Economics, 2016. — P. 13–38.
- [31] Vinogradov Victor V. The history of words: About 1500 words and expressions and more than 5000 words associated with them [Istoriia slov: Ok. 1500 slov i vyrazhenii i bolee 5000 slov, s nimi sviaz. — Russian language [Russkii yazyk], 1999.