UWB@RuShiftEval Measuring Semantic Difference as per-word Variation in Aligned Semantic Spaces

Přibáň, P.[* 1,2], **Pražák, O.**[* 1,2], and **Taylor, S. E.**[* 2]

[1]NTIS – New Technologies for the Information Society,
[2]Department of Computer Science and Engineering,
Faculty of Applied Sciences, University of West Bohemia, Czech Republic
{pribanp, ondfa, taylor}@kiv.zcu.cz
http://nlp.kiv.zcu.cz

Abstract

We present our system for measuring the lexical semantic change between corpora pairs, developed for the *RuShiftEval* competition. We measure semantic changes in a list of test words between three different corpora (for each pair), *pre-Soviet*, *Soviet*, and *post-Soviet*, each from a different time period. For each corpus, we train word embeddings, and we obtain linear transformations between the embedding spaces. Finally, we measure the similarity between the transformed vectors for each test word.

Keywords: Diachronic Semantic Change, Word Embeddings, Aligning Word Embeddings, Linear Transformations between Semantic Spaces

DOI: 10.28995/2075-7182-2021-20-XX-XX

UWB@RuShiftEval Измерение семантической разницы как вариации для каждого слова в выровненных семантических пространствах

Пршибан, П. (pribanp@kiv.zcu.cz), Пражак, О. (ondfa@kiv.zcu.cz),
Тейлор, С. Э. (taylor@kiv.zcu.cz)

Университет Западной Богемии, Пльзень, Чехия

1 Introduction

Words have a history; that is, they change and evolve over time. Types of changes include: orthographic, or changes in spelling or capitalization; register changes, in which a word becomes more respectable or less; and the semantic changes discussed in this paper, in which the word changes its relationships with other words in the vocabulary, perhaps by losing or gaining a sense or just changing the frequency of use for a sense.

The task of finding word sense changes over time belongs to the field of natural language processing. It is called diachronic *Lexical Semantic Change* (LSC) and in recent years, it is getting more attention [15, 12, 29, 6]. There is also the *synchronic LSC* task, which aims to identify domain-specific changes of word senses compared to general-language usage [35].

1.1 The RuShiftEval Task

The goal of the RuShiftEval task[1] [18] is to rank a set of target words according to their degree of lexical semantic similarity between two corpora C_i and C_j, each from a different time period t_i and t_j, respectively. A lower rank (score) for the test word means stronger semantic change. The organizers

[*]Equal contribution.

[1]It is almost the same task as the *SemEval-2020 Task 1*, sub-task 2, see Section 2.

provided three corpora for different time periods C_1 *pre-Soviet* (1700-1916), C_2 *Soviet* (1918-1990) and C_3 *post-Soviet* (1991-2016), see Section 3 for details. They also provided annotated training data (ranks for list of words) for two corpora pairs C_1/C_2 (*pre-Soviet/Soviet*) and C_2/C_3 (*Soviet/post-Soviet*) from the RuSemShift test set [25]. In the test phase, the participants were asked to produce outputs for a list of test words[2] for all three pairs of the corpora, i.e., C_1/C_2 (*pre-Soviet/Soviet*), C_2/C_3 (*Soviet/post-Soviet*) and C_1/C_3 (*pre-Soviet/post-Soviet*). In evaluation, a Spearman correlation coefficient with gold rankings based on human annotation is computed for each pair of corpora and subsequently all three coefficients are averaged. The average is used as a final score for one submission.

Even though the organizers provided some annotated training data, we did not use the data and we rely on a completely unsupervised solution. We use a very similar approach to [34, 23]. The general idea of our solution is that we consider each pair of different corpora C_i and C_j as separate languages L_1 and L_2 despite the fact that both of them are written in Russian. We assume that both of these *languages* are deeply similar in all aspects, including semantic. We train a semantic space (word embedding) for each corpus and for each pair of corpora, we perform a cross-lingual mapping (i.e., we transform the spaces into one common space) of their corresponding word embeddings. For the cross-lingual transformation, we use the Orthogonal transformation [1, 2, 11, 3] and Canonical Correlation Analysis (CCA) [5]. After the transformation for one corpora pair, we can measure cosine similarity for one target word between its two vectors because they are in the same space. The cosine similarity denotes their semantic similarity, all the target words are then ranked according to the cosine similarity.

2 Related Work

Word embeddings are vector spaces in which the location in the space reveals some relationships with nearby words. For example, the nearest-neighbor set of a word generally includes words which have some semantic or syntactic relationship with it. The embedding is thus a concrete demonstration of Harris' remark *difference of meaning correlates with difference of distribution* [16].

Alignment transformations or *cross-lingual transformations* are techniques for aligning two word embedding spaces, first described in Mikolov et al. [20] for translating single words from one language to another. For example, to find English equivalents of Russian words, semantic word embedding vector spaces X_E and X_R with vector length k for each language are built from monolingual corpora C_E, C_R. A small dictionary of cross-language synonyms, with n pairs like e.g. (Please, Просим) is provided. These synonyms are used to create two n by k matrices D_E and D_R, with rows in the D_E matrix consisting of semantic vectors from the English embedding, and corresponding rows in the D_R matrix consisting of semantic vectors from the Russian embedding. Finally a linear transformation $W^{E \to R}$ is trained to change the vectors for the English vectors to the Russian ones, by finding the following minimum:

$$W^{E \to R} = \underset{W}{\operatorname{argmin}} \; ||D_R - WD_R||^2$$

Then we hope to find that vector v_q^E for an English word q not in our dictionary will be mapped with $W^{E \to R} v_q^E$ to be in the near neighborhood of vectors $v_{w_j}^R$, good Russian translations of q.

Kulkarni et al. [30] may have been the first to use the displacement of a word mapped by a linear transformation from one word-embedding semantic space to another as a measurement of semantic change. Their transformations, between word embeddings for corpora C_s and C_t with large vocabulary intersections, were created based on nearest neighbors. For each test word q, if $v_{w_i}^s$ is the vector for the i-th nearest neighbor to v_q^s in one space, and $v_{w_i}^t$ is the vector for the same word in the other space, and each such vector is k elements long, then a $k \times k$ matrix $W^{s \to t}$ such that $v_{w_i}^t = W^{s \to t} v_{w_i}^s$ is completely determined whenever we consider at least k nearest neighbors, that is $|v_{w_i}^s| > k$. Thus, they build a piece-wise linear transformation from one space to another around each test word, and compare the mapped position of the test word to the actual position for the same word in the second space.

Hamilton et al. [15] used the term *orthogonal Procrustes*[3] to describe their method of building an

[2] The list was same for all three pairs.

[3] Procrustes was a figure in Greek mythology who forced passing travellers to fit into his bed, either stretching them longer or cutting them shorter if necessary.

orthogonal linear mapping to align embeddings for two time periods. Like Mikolov [20] and unlike Kulkarni, they use one transformation for all words in the vocabulary. Unlike Mikolov, they add the additional constraint that the transformation be orthogonal, so that angles and thus cosine similarities between words are preserved by the transformation. With this added condition, the problem has a closed form using Singular Value Decomposition. Their experiments showed that the combination of word-embeddings and orthogonal Procrustes give excellent results when there is sufficient data to build good word embeddings.

Tahmasebi et al. [32] provide a comprehensive survey of techniques for the LSC task. Schlechtweg et al. [35] evaluate available approaches for *LSC* detection using the *DURel* dataset [28].

According to [35], there are three main types of approaches. (1) Semantic vector spaces approaches [13, 8, 14, 15, 26, 34, 23] represent each word with two vectors for two different time periods. The change of meaning is then measured by some distance (usually by the cosine distance) between the two vectors. (2) Topic modeling approaches [4, 19, 21, 10, 27] estimate a probability distribution of words over their different senses, i.e., topics and (3) Clustering models [36, 33].

The recent competitions focused on LSC: *SemEval-2020 Task 1* [29] and *DIACR-Ita* [6] focused on unsupervised approaches and did not provide any annotated training data. The goal of the *SemEval-2020 Task 1* was to measure the binary (changed or not) and continuous change in words between two time periods in English, German, Latin and Swedish corpora. In the *DIACR-Ita* the participants were asked only to measure the binary changes for a set of target words between two time periods in Italian.

3 Data

The data for the RuShiftEval task is drawn from the Russian National Corpus[4]. Like the data for the other tasks shown below, each corpus consists of random sentences chosen from the literature of its time-period. Thus, neighboring sentences provide no additional context for words in short sentences.

Table 1: Comparing corpus token count and test word count on recent tasks.

Task:Language	Period 1	Period 2	Period 3	# test words
SemEval T1:English	6.5M	6.7M	–	37
SemEval T1:German	70.2M	72.3M	–	48
SemEval T1:Latin	1.7M	9.4M	–	40
SemEval T1:Swedish	71M	110.7M	–	31
DIACR-Ita:Italian	156.8M	589.6M	–	18
RuShiftEval:Russian	75.1M	97.2M	85.3M	99

The RuShiftEval corpora are not lemmatized; the tokens in the corpora are inflected wordforms used correctly in the sentence context. The test words, however, are lemmas. Some of them may appear in the corpora when an inflected word is the same as the lemma, but in general, many of the uses of a lemma are different than the lemma itself.

For the *SemEval-2020 Task 1*, the corpora were provided only in lemmatized form, in part because of copyright issues. Reading them, it is clear that some information has been lost, including number and case for nouns and person and tense for verbs. These *might* not be important to the task, but there is no choice to make; all the evaluations must use the lemmatized forms (which are not *always* the actual lemmas; the text processing has a few errors). For *DIACR-Ita*, the corpus was provided in vertical form, with the original token, the lemma, and the part of speech available for each token. Prazak et al. [23] considered four of the seven possible choices: tokens; token_POS (Part-of-Speech); lemmas; and lemma_POS. The change measurements were most consistent for lemmas, and least consistent for token_POS, with the other two choices in the middle. For most of our RuShiftEval submissions, we chose to use lemmas instead of raw tokens.

We use UDpipe [31] with the `russian-syntagrus-ud-2.5-191206.udpipe` model to lemmatize the corpora, that is, to convert every token to its lemma.

[4]https://rusvectores.org/static/corpora/

4 System Description

Our method[5] is fully unsupervised and language-independent. It consists of preparing a semantic vector space for each corpus, earlier and later; computing a linear transformation between earlier and later spaces, using Canonical Correlation Analysis and Orthogonal Transformation; and measuring the cosines between the transformed vector for the target word from the earlier corpus and the vector for the target word in the later corpus.

First, we train semantic spaces from each corpus C_1, C_2 and C_3. We represent the semantic spaces by a matrix $\mathbf{X}^s$ (i.e., a source space s) and a matrix $\mathbf{X}^t$ (i.e, a target space t) using fastText [9] and word2vec [7] Skip-gram with negative sampling. We perform a cross-lingual mapping of the two vector spaces, getting two matrices $\hat{\mathbf{X}}^s$ and $\hat{\mathbf{X}}^t$ projected into a shared space. We select two methods for the cross-lingual mapping *Canonical Correlation Analysis (CCA)* using our own implementation and a modification of the *Orthogonal Transformation* from *VecMap* [3]. Both of these methods are linear transformations. In our case, the transformation can be written as follows:

$$\hat{\mathbf{X}}^s = \mathbf{W}^{s \to t} \mathbf{X}^s \tag{1}$$

where $\mathbf{W}^{s \to t}$ is a matrix that performs linear transformation from the source space s (matrix $\mathbf{X}^s$) into a target space t and $\hat{\mathbf{X}}^s$ is the source space transformed into the target space t (the matrix $\mathbf{X}^t$ does not have to be transformed because $\mathbf{X}^t$ is already in the target space t and $\mathbf{X}^t = \hat{\mathbf{X}}^t$).

Generally, the CCA transformation transforms both spaces $\mathbf{X}^s$ and $\mathbf{X}^t$ into a third shared space o (where $\mathbf{X}^s \neq \hat{\mathbf{X}}^s$ and $\mathbf{X}^t \neq \hat{\mathbf{X}}^t$). Thus, CCA computes two transformation matrices $\mathbf{W}^{s \to o}$ for the source space and $\mathbf{W}^{t \to o}$ for the target space. The transformation matrices are computed by minimizing the negative correlation between the vectors $\mathbf{x}_i^s \in \mathbf{X}^s$ and $\mathbf{x}_i^t \in \mathbf{X}^t$ that are projected into the shared space o. The negative correlation ρ is defined as follows:

$$\underset{\mathbf{W}^{s \to o}, \mathbf{W}^{t \to o}}{\operatorname{argmin}} \; -\sum_{i=1}^{n} \rho(\mathbf{W}^{s \to o}\mathbf{x}_i^s, \mathbf{W}^{t \to o}\mathbf{x}_i^t) = -\sum_{i=1}^{n} \frac{cov(\mathbf{W}^{s \to o}\mathbf{x}_i^s, \mathbf{W}^{t \to o}\mathbf{x}_i^t)}{\sqrt{var(\mathbf{W}^{s \to o}\mathbf{x}_i^s) \times var(\mathbf{W}^{t \to o}\mathbf{x}_i^t)}} \tag{2}$$

where cov the covariance, var is the variance and n is a number of vectors. In our implementation of CCA, the matrix $\hat{\mathbf{X}}^t$ is equal to the matrix $\mathbf{X}^t$ because it transforms only the source space s (matrix $\mathbf{X}^s$) into the target space t from the common shared space with a pseudo-inversion, and the target space does not change. The matrix $\mathbf{W}^{s \to t}$ for this transformation is then given by:

$$\mathbf{W}^{s \to t} = \mathbf{W}^{s \to o}(\mathbf{W}^{t \to o})^{-1} \tag{3}$$

In the case of the *Orthogonal Transformation*, the submission is referred to as *ort-200-OT*, see Table 2. We use Orthogonal Transformation with a supervised seed dictionary consisting of all words common to both semantic spaces. The transformation matrix $\mathbf{W}^{s \to t}$ is given by:

$$\underset{\mathbf{W}^{s \to t}}{\operatorname{argmin}} \sum_{i}^{|V|} (\mathbf{W}^{s \to t}\mathbf{x}_i^s - \mathbf{x}_i^t)^2 \tag{4}$$

under the hard condition that $\mathbf{W}^{s \to t}$ needs to be orthogonal, where V is the vocabulary of correct word translations from source to target space. The reason for the orthogonality constraint is that linear transformation with the orthogonal matrix does not squeeze or re-scale the transformed space. It only rotates the space, thus it preserves most of the relationships of its elements (in our case, it is important that orthogonal transformation preserves angles between the words, so it preserves the cosine similarity).

Finally, in all transformation methods, for each word w_i from the set of target words T, we select its corresponding vectors $\mathbf{v}_{w_i}^s$ and $\mathbf{v}_{w_i}^t$ from matrices $\hat{\mathbf{X}}^s$ and $\hat{\mathbf{X}}^t$, respectively ($\mathbf{v}_{w_i}^s \in \hat{\mathbf{X}}^s$ and $\mathbf{v}_{w_i}^t \in \hat{\mathbf{X}}^t$), and we compute the cosine similarity between these two vectors. The cosine similarity is then used to generate an output, i.e., the ranking.

[5]The source code is available at `https://github.com/pauli31/SemEval2020-task1`

Table 2: Results of our selected submissions for all three corpora pairs.

Rank	C_1/C_2	C_2/C_3	C_1/C_3	Avg.	Description
27	.367	.354	.533	.417	cca-150-OT
49	.277	.273	.464	.338	cca-100to220-agg
50	.239	.307	.450	.332	ort-200-OT
57	.220	.255	.446	.307	cca-100to500-agg
71	.096	.155	.317	.190	cca-300-pre-/post-trained

Aggregating runs (submissions ranked 49 and 57) were performed by giving each test word the average of its rank in the individual runs. Our best aggregating run is *cca-100to220-agg*, aggregating runs from 29 embeddings with vector lengths from 100 to 220. *Cca-100to500-agg* aggregates runs on 119 embeddings with vector lengths from 100 to 500. It has a smaller Standard Error of the Mean, but worse scores.

Pre-training (submission ranked 71 in Table 2) was performed for one epoch on the entire vocabulary for all corpora, and then for five epochs in one corpus. The pre-training was supposed to start all three embeddings from the same random beginning. For training our embeddings, including pre-training we used the gensim [24] python module, which re-implements both the word2vec Skip-gram and the fastText training algorithms. Both aggregation and pre-training submissions use the CCA transformation.

For the experiments with the orthogonal transformation, we use the *VecMap* tool [3]. We use embeddings with the size of 200 and 300. We performed two versions of the experiments, the first with lemmatized text and the other with only the target word lemmatized and the rest of the text only lowercased (*OT* in the name of the submissions).

The *cca-150-OT* submission used the CCA transformation with fastText embeddings with dimension 150. We build the translation dictionary for the transformation of the two spaces by removing the target words from the intersection of their vocabularies (we use all other words from the intersection). The embeddings are trained on the same preprocessed text as the orthogonal submission.

4.1 Results

We made ten submissions. The most interesting submissions are shown in table 2, which is excerpted from the full leaderboard. Our team ranked near the median, with absolute correlations less than we expected based on our scores with RuSemShift data.

As described above, we experimented with several different details within the broad strategy of aligning embeddings. Although single runs on each sub-strategy are not necessarily statistically significant, table 2 shows that the OT (only targets lemmatized) sub-strategy and CCA transformations were relatively successful, and that embeddings with more than 220-dimensional vectors were less successful, perhaps because more training should have been used for the extra parameters.

5 Conclusion

In the two earlier recent semantic change workshops [29, 6], strategies like ours did very well [22, 17]. In contrast, at RuShiftEval, other strategies (probably supervised) have dramatically taken over the top spots. You will have to read those papers to find out how they did it. However, we can speculate on some differences peculiar to this situation.

Our approach is unsupervised. We did a couple of sanity runs using the RuSemShift data [25], but made no effort to exploit it. So systems which found ways to incorporate supervision *should* do better.

It is possible that the human annotation of the gold data is capturing only a part of the total semantic change in test words, and thus supervised systems which attempt to match those annotations fare better than those like ours which look at global usage. Against this theory, the annotation technique appears to be very similar to that of Schlechtweg et al. [28], used in the other recent workshops.

None of the languages for the other workshops in Table 1 is from the Slavic language group. Slavic languages have more cases for nouns, and perhaps more tokens/lemma than other languages. They have a generally freer word-order, which may have an effect on a five-word context window. Latin has some of the same problems as the Slavic languages, and Latin results for the SemEval task were generally much worse – but the small corpora sizes for Latin could have aggravated those problems (see Table 1), and were certainly easy to blame. In contrast, the RuShiftEval corpora are completely adequate.

Further, because it is easier to get a good Spearman correlation for a shorter sequence, our 'sanity' runs on the RuSemShift data may have misled us about our likely scores on the final evaluation. The expected absolute value of the Spearman correlation of two lists of random numbers is 0.27 for lists of length 10, 0.11 for lists of length 50, and 0.08 for lists of length 100. The shorter lists also have larger standard deviation. The lists of test words for previous workshops ranged from 18 to 48 words long, and the two RuSemShift lists are 49 and 50 words. So higher scores are easier to achieve for those shorter lists. The highest score of ten submissions is most likely to be 1.3σ above the mean. (Another factor is averaging the three inter-corpora scores. Assuming that the standard deviation of each score is s, the standard deviation of the average is $\frac{s}{\sqrt{3}}$.) Everyone had this problem; but it may partly account for our confusion about how well we expected to do.

Finally, of course, we may have made a silly implementation error. If so, we were in good company.

6 Acknowledgements

This work has been partly supported by ERDF "Research and Development of Intelligent Components of Advanced Technologies for the Pilsen Metropolitan Area (InteCom)" (no.: CZ.02.1.01/0.0/0.0/17 048/0007267); by the project LO1506 of the Czech Ministry of Education, Youth and Sports; and by Grant No. SGS-2019-018 Processing of heterogeneous data and its specialized applications. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

References

[1] Artetxe Mikel, Labaka Gorka, Agirre Eneko. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. — Austin, Texas : Association for Computational Linguistics, 2016. — November. — P. 2289–2294. — Access mode: https://aclweb.org/anthology/D16-1250.

[2] Artetxe Mikel, Labaka Gorka, Agirre Eneko. Learning bilingual word embeddings with (almost) no bilingual data // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). — Vancouver, Canada : Association for Computational Linguistics, 2017. — Jul. — P. 451–462. — Access mode: https://www.aclweb.org/anthology/P17-1042.

[3] Artetxe Mikel, Labaka Gorka, Agirre Eneko. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). — Melbourne, Australia : Association for Computational Linguistics, 2018. — Jul. — P. 789–798. — Access mode: https://www.aclweb.org/anthology/P18-1073.

[4] Bamman David, Crane Gregory. Measuring Historical Word Sense Variation // Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries. — JCDL '11. — New York, NY, USA : Association for Computing Machinery, 2011. — P. 1–10. — Access mode: https://doi.org/10.1145/1998076.1998078.

[5] Brychcín Tomáš, Taylor Stephen, Svoboda Lukáš. Cross-lingual word analogies using linear transformations between semantic spaces // Expert Systems with Applications. — 2019. — Vol. 135. — P. 287–295.

[6] DIACR-Ita @ EVALITA2020: Overview of the EVALITA2020 Diachronic Lexical Semantics (DIACR-Ita) Task / Pierpaolo Basile, Annalina Caputo, Tommaso Caselli et al. // Proceedings of

the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020) / Ed. by Valerio Basile, Danilo Croce, Maria Di Maro, Lucia C. Passaro. — Online : CEUR.org, 2020.

[7] Efficient Estimation of Word Representations in Vector Space / Tomás Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean // 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings / Ed. by Yoshua Bengio, Yann LeCun. — 2013. — Access mode: `http://arxiv.org/abs/1301.3781`.

[8] Eger Steffen, Mehler Alexander. On the Linearity of Semantic Change: Investigating Meaning Variation via Dynamic Graph Models // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). — Berlin, Germany : Association for Computational Linguistics, 2016. — Aug. — P. 52–58. — Access mode: `https://www.aclweb.org/anthology/P16-2009`.

[9] Enriching Word Vectors with Subword Information / Piotr Bojanowski, Edouard Grave, Armand Joulin, Tomas Mikolov // arXiv preprint arXiv:1607.04606. — 2016.

[10] Fremann Lea, Lapata Mirella. A Bayesian Model of Diachronic Meaning Change // Transactions of the Association for Computational Linguistics. — 2016. — Vol. 4. — P. 31–45. — Access mode: `https://www.aclweb.org/anthology/Q16-1003`.

[11] Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations / Mikel Artetxe, Gorka Labaka, , Eneko Agirre // Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). — 2018. — P. 5012–5019. — Access mode: `https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16935/16781`.

[12] German in Flux: Detecting Metaphoric Change via Word Entropy / Dominik Schlechtweg, Stefanie Eckmann, Enrico Santus et al. // Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017). — Vancouver, Canada : Association for Computational Linguistics, 2017. — Aug. — P. 354–367. — Access mode: `https://www.aclweb.org/anthology/K17-1036`.

[13] Gulordava Kristina, Baroni Marco. A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. // Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics. — Edinburgh, UK : Association for Computational Linguistics, 2011. — Jul. — P. 67–71. — Access mode: `https://www.aclweb.org/anthology/W11-2508`.

[14] Hamilton William L., Leskovec Jure, Jurafsky Dan. Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. — Austin, Texas : Association for Computational Linguistics, 2016. — Nov. — P. 2116–2121. — Access mode: `https://www.aclweb.org/anthology/D16-1229`.

[15] 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). — Berlin, Germany : Association for Computational Linguistics, 2016. — Aug. — P. 1489–1501. — Access mode: `https://www.aclweb.org/anthology/P16-1141`.

[16] Harris Zellig S. Distributional structure // Word. — 1954. — Vol. 10, no. 23. — P. 146–162.

[17] Kaiser Jens, Schlechtweg Dominik, im Walde Sabine Schulte. OP-IMS @ DIACR-Ita: Back to the Roots: SGNS+OP+CD still Rocks Semantic Change Detection // Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online event, December 17th, 2020 / Ed. by Valerio Basile, Danilo Croce,

Maria Di Maro, Lucia C. Passaro. — Vol. 2765 of CEUR Workshop Proceedings. — CEUR-WS.org, 2020. — Access mode: http://ceur-ws.org/Vol-2765/paper133.pdf.

[18] Kutuzov Andrey, Pivovarova Lidia. RuShiftEval: a shared task on semantic shift detection for Russian // Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference. — 2021.

[19] Mihalcea Rada, Nastase Vivi. Word Epoch Disambiguation: Finding How Words Change Over Time // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). — Jeju Island, Korea : Association for Computational Linguistics, 2012. — Jul. — P. 259–263. — Access mode: https://www.aclweb.org/anthology/P12-2051.

[20] Mikolov Tomas, Le Quoc V., Sutskever Ilya. Exploiting Similarities among Languages for Machine Translation // arXiv. — 2013. — 1309.4168.

[21] Novel Word-sense Identification / Paul Cook, Jey Han Lau, Diana McCarthy, Timothy Baldwin // Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. — Dublin, Ireland : Dublin City University and Association for Computational Linguistics, 2014. — Aug. — P. 1624–1635. — Access mode: https://www.aclweb.org/anthology/C14-1154.

[22] Pömsl Martin, Lyapin Roman. CIRCE at SemEval-2020 Task 1: Ensembling Context-Free and Context-Dependent Word Representations // Proceedings of the Fourteenth Workshop on Semantic Evaluation. — Barcelona (online) : International Committee for Computational Linguistics, 2020. — Dec. — Access mode: https://www.aclweb.org/anthology/2020.semeval-1.21.pdf.

[23] Pražák Ondrej, Přibáň Pavel, Taylor Stephen. UWB @ DIACR-Ita: Lexical Semantic Change Detection with CCA and Orthogonal Transformation // Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online event, December 17th, 2020 / Ed. by Valerio Basile, Danilo Croce, Maria Di Maro, Lucia C. Passaro. — Vol. 2765 of CEUR Workshop Proceedings. — CEUR-WS.org, 2020. — Access mode: http://ceur-ws.org/Vol-2765/paper110.pdf.

[24] Řehůřek Radim, Sojka Petr. Software Framework for Topic Modelling with Large Corpora // Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. — Valletta, Malta : ELRA, 2010. — May. — P. 45–50.

[25] Rodina Julia, Kutuzov Andrey. RuSemShift: a dataset of historical lexical semantic change in Russian // Proceedings of the 28th International Conference on Computational Linguistics. — Barcelona, Spain (Online) : International Committee on Computational Linguistics, 2020. — Dec. — P. 1037–1047. — Access mode: https://www.aclweb.org/anthology/2020.coling-main.90.

[26] Rosenfeld Alex, Erk Katrin. Deep Neural Models of Semantic Shift // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). — New Orleans, Louisiana : Association for Computational Linguistics, 2018. — Jun. — P. 474–484. — Access mode: https://www.aclweb.org/anthology/N18-1044.

[27] Schlechtweg Dominik, Walde Sabine Schulte im. Simulating Lexical Semantic Change From Sense-Annotated Data // The Evolution of Language: Proceedings of the 13th International Conference (EvoLang13) / Ed. by A. Ravignani, C. Barbieri, M. Martins et al. — 2020. — Access mode: http://brussels.evolang.org/proceedings/paper.html?nr=9.

[28] Schlechtweg Domininк, im Wlade Sabine Schulte, Eckmann Stefanie. Diachronic Usage Relatedness (DURel): A Framework for the Annotation of Lexical Semantic Change // Proceedings of NAACL-HLT 2018. — 2018. — P. 169–174.

[29] SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection / Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen et al. // Proceedings of the Fourteenth Workshop on Semantic Evaluation. — Barcelona (online) : International Committee for Computational Linguistics,

2020. — Dec. — P. 1–23. — Access mode: https://www.aclweb.org/anthology/2020.semeval-1.1.

[30] Statistically Significant Detection of Linguistic Change / Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, Steven Skiena // Proceedings of the 24th International Conference on World Wide Web. — 2015. — P. 625–635. — Access mode: http://www.perozzi.net/publications/15_www_linguistic.pdf.

[31] Straka Milan, Straková Jana. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe // Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. — Vancouver, Canada : Association for Computational Linguistics, 2017. — August. — P. 88–99. — Access mode: http://www.aclweb.org/anthology/K/K17/K17-3009.pdf.

[32] Tahmasebi Nina, Borin Lars, Jatowt Adam. Survey of computational approaches to lexical semantic change // arXiv preprint arXiv:1811.06278. — 2018.

[33] Tahmasebi Nina, Risse Thomas. Finding Individual Word Sense Changes and their Delay in Appearance // Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017. — Varna, Bulgaria : INCOMA Ltd., 2017. — Sep. — P. 741–749. — Access mode: https://doi.org/10.26615/978-954-452-049-6_095.

[34] UWB at SemEval-2020 Task 1: Lexical Semantic Change Detection / Ondřej Pražák, Pavel Přibáň, Stephen Taylor, Jakub Sido // Proceedings of the Fourteenth Workshop on Semantic Evaluation. — Barcelona (online) : International Committee for Computational Linguistics, 2020. — Dec. — P. 246–254. — Access mode: https://www.aclweb.org/anthology/2020.semeval-1.30.

[35] A Wind of Change: Detecting and Evaluating Lexical Semantic Change across Times and Domains / Dominik Schlechtweg, Anna Hätty, Marco Del Tredici, Sabine Schulte im Walde // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. — Florence, Italy : Association for Computational Linguistics, 2019. — Jul. — P. 732–746. — Access mode: https://www.aclweb.org/anthology/P19-1072.

[36] An automatic approach to identify word sense changes in text media across timescales / Sunny Mitra, Ritwik Mitra, Suman Kalyan Maity et al. // Natural Language Engineering. — 2015. — Vol. 21, no. 5. — P. 773–798.