Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2019"

Moscow, May 29—June 1, 2019

WORD VECTOR MODELS AS AN OBJECT OF LINGUISTIC RESEARCH

Shavrina T. O. (rybolos@gmail.com)

NRU HSE, Moscow, Russia; Sberbank, Moscow, Russia

This article launches a series of studies in which popular vector word2vec models are considered not as an element of the architecture of an NLP application, but as an independent object of linguistic research. The linguist's view on the surrogate of contexts on the corpus, as which vector models can be considered, makes it possible to reveal new information about the distribution of individual semantic groups of vocabulary and new knowledge about the corpus from which these models are derived. In particular, it is shown that such layers of English and Russian vocabulary, such as the names of professions, nationalities, toponyms, personal qualities, time periods, have the greatest independence from changing the model and retain their position relative to their neighbour words—that is, they have the most stable contexts regardless of the corpus; it is shown that the vocabulary from the Swadesh list is statistically more resistant to changing the model than the frequency vocabulary is; it is shown which word2vec models for the Russian language preserve best the ontological structures in vocabulary.

Key words: word2vec, word vector model, word vectors, vector model evaluation, word2vec interpretation

ВЕКТОРНЫЕ МОДЕЛИ КАК ОБЪЕКТ ЛИНГВИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

Шаврина Т. О. (rybolos@gmail.com)

НИУ ВШЭ, Москва, Россия; Сбербанк, Москва, Россия

В данной статье начата серия исследований, в которых популярные векторные модели word2vec рассматриваются не как элемент архитектуры NLP-приложения, а как самостоятельный объект лингвистического исследования. Взгляд лингвиста на суррогат контекстов, коим можно назвать такие модели, позволяет выявить новую информацию о распределении отдельных семантических групп лексики и о корпусах, на которых эти модели получены. В частности, показывается, что такие пласты английской и русской лексики, как названия профессий, национальностей, топонимы, качества личности, временные сроки, обладают наибольшей независимостью от смены модели и сохраняют свое положение относительно соседей — то есть имеют наиболее устойчивые контексты независимо от корпуса; показывается, что лексика из списка Сводеша в среднем более устойчива к смене модели, чем частотная лексика; показывается, какие модели word2vec для русского языка наилучшим образом сохраняют онтологические структуры в лексике.

Ключевые слова: word2vec, векторные модели, word vectors, evaluation

1. About word vector models

Vector word models are currently one of the main elements in architectures for language modelling and processing, showing themselves to be an effective way to convey information about the meaning and generalized contexts of individual words. From the point of view of mathematics, modern vector skip-gram and CBOW models have an indisputable advantage over other ways of vectorizing words—they simultaneously describe the distribution of words relative to each other and also take into account their sequential order.

However, the assessment of such vector models like word2vec [Mikolov et al., 2013], GloVE [Pennington et al., 2014], fasttext [Bojanowski et al., 2017] is currently hampered by the "black box" of the algorithm for obtaining them—and the quality of the models is estimated very indirectly. This study is devoted to the development of a linguistic apparatus for assessing the quality of vector models based on linguistic knowledge.

The main hypothesis on which the training word2vec models is based is "the words having the same contexts mean the same". Both Skip-gram and CBOW models provide high-quality word embeddings with this hypothesis, however, any linguist will call the resulting problems:

- there are words with similar contexts meaning the opposite—antonyms;
- there are words with different contexts meaning the same—historical synonyms, multi-word expressions, etc;
- also, well-known problems are polysemy, morphological derivatives, misprints.

These problems lead to the attention shift in the human evaluation of vector models: the vocabulary of the medium frequency, non-homonymous, unambiguous gives those beautiful examples of vector calculations (*"king"—'man' = 'queen', etc. by* [Mikolov et al., 2013]). What happens on the other groups of lexis?

For a linguistic point of view, the word vector model is a linguistic surrogate all the contexts in a corpus from which it is derived. Thus it can be considered interesting as an independent object of study, object situated in the middle of the usual division [Saussure, 1916] of synchronic and diachronic approaches: cumulative information about word behaviour in the language, obtained on the basis of all contexts over a certain (usually broad) time period, can be surprisingly accurate—examples (1) and (2) show that such a model can even accumulate extralinguistic knowledge if trained on the billionth volume of words.

(1) *5 closest words to the word 'otradnoye'* (adj, name of a Moscow metro station) *on word2vec model are 5 geographically adjacent Metro stations* (model trained on Russian National Corpus).

Semantic associates for отрадное[1] (computed on Ruscorpora and Wikipedia)

лобаново	*0.614*
петровско-разумовское	*0.585*
романово	*0.577*
глухово	*0.574*
лукино	*0.572*

(2) *10 closest words to the word 'shabolovskaya'* (adj, name of a station on crossing Metro lines) *on word2vec model are geographically adjacent Metro stations, street names and names of crossing metro lines* (model trained on news corpus).

Semantic associates for шаболовская[2] (ADJ) computed on news corpus

шаболовский$_{\text{ADJ}}$	*0.59*
щёлковская$_{\text{ADJ}}$	*0.53*
серпуховской$_{\text{ADJ}}$	*0.51*
-радиальный$_{\text{ADJ}}$	*0.49*
таганско$_{\text{ADJ}}$	*0.49*
198-ть$_{\text{ADJ}}$	*0.49*
добрынинская$_{\text{ADJ}}$	*0.49*
филетовый$_{\text{ADJ}}$	*0.48*
подбельский$_{\text{ADJ}}$	*0.47*
калужско-рижский$_{\text{ADJ}}$	*0.47*

Hereinafter, all results will be presented on RusVectores project [Kutuzov, Andreev, 2015], [Kutuzov, Kuzmenko, 2017] models—all skip-gram, with lemmatization and pos-tagging, trained on 1) news corpus[3], 2) Russian National Corpus[4] and Wikipedia, 3) Taiga corpus[5] and 4) Aranea corpus[6]. Results in English are computed on a sister project—WebVectors[7]

---

1 https://rusvectores.org/en/ruwikiruscorpora_upos_skipgram_300_2_2019/отрадное_ADJ/

2 https://rusvectores.org/en/news_upos_skipgram_300_5_2019/шаболовская_ADJ/

3 News: news stream from 1500 primarily Russian-language news sites, model: http://vectors.nlpl.eu/repository/11/184.zip

4 Full Russian National Corpus http://ruscorpora.ru/en/, model https://rusvectores.org/static/models/rusvectores4/RNC/ruscorpora_upos_skipgram_300_5_2018.vec.gz

5 **Taiga:** open and structured Russian web corpus https://tatianashavrina.github.io/taiga_site/, model https://rusvectores.org/static/models/rusvectores4/taiga/taiga_upos_skipgram_300_2_2018.vec.gz

6 **Araneum Russicum Maximum:** large web corpus of Russian http://ella.juls.savba.sk/aranea_about, model https://rusvectores.org/static/models/rusvectores4/araneum/araneum_upos_skipgram_300_2_2018.vec.gz

7 http://vectors.nlpl.eu/explore/embeddings/en/about/

2. Word vector model evaluation

Vector models are of great interest in connection with the mediated material they represent—for the needs of corpora comparison and assessment, for analyzing the nature of lexis. Knowledge of the "normal" and "anomalous" behaviour of lexis on the corpora would allow a much more accurate assessment of the quality of the obtained model vectors.

However, the quality assessment of vector models is still fairly superficial—this is either enumerating all possible models and choosing one that showed the best result in a particular architecture and specific task [Kutuzov, 2015], or an assessment on a small set of individual pairs of words with human assessment of their distance (completely subjective)—SimLex999 [Hill et al. 2015] and Google Analogy [Mikolov 2013]. Several significant studies [Tsvetkov et al. 2016], [Vulich et al. 2017] have already shown that the quality of vector models for the English language is unstable and depends on many factors, and for an independent assessment of models, a new methodological apparatus is needed.

The evaluation problem grows like a snowball—in 2018, the first studies devoted to obtaining the best vector models were published, claiming universality for all words and sentences in a language—BERT [Devlin et al. 2018], ELMo [Peters et al. 2018], and OpenAI architecture [Radford et al. 2018]. The main trend in NLP remains—we search for an effective way to vectorize words and whole texts, but to evaluate model effectiveness, a new approach and a new level of understanding of the resulting models despite the corpus features is missing. Next, we consider a series of experiments devoted to the study of the lexis behaviour in word2vec models and the linguistic interpretation of the quality of word vectors—the preservation of known ontological relationships, most stable vocabulary groups, and so on.

3. The behaviour of lexis in word vector models

In accordance with the first hypothesis about the lexis behaviour in word2vec models, it was decided to check the Swadesh list [Swadesh, 1950]—words from a manually compiled list that are considered chronologically the most stable in the language. Words from the Swadesh list do have interesting characteristics from the point of view of vector models—they denote the basic concept—relatives, animals, main action verbs, colorus, numbers, etc., and have a frequency above the average, that is, have enough contexts in any corpus. Hypothetically, on vector models, such vocabulary should be stable relative to its neighbours.

3.1. Experiment 1

Swadesh list was obtained for Russian and English in its fullest form (200 words), then only those words that were found on all models in concern were left—these are 173 words for Russian and 160 words for English since stop words are removed from the models before training[8].

[8] The full list can be found in the repository https://github.com/TatianaShavrina/wordvector_metrics.

Then, for each of the words in the list, the share of the word neighbours always presented regardless of the model was calculated—in the window of the 10 closest ones, as well as the 20, 50, 100, 200 and 300 nearest neighbours. For the Russian language, the models RNC + wiki, Taiga, Aranea were used, and for English—BNC, Wiki, Gigaword.

For comparison, random words of a general dictionary of models were also taken, and, separately, random words with a high frequency (top 2000). For the Russian language, frequencies were taken from [Lyashevskaya, Sharoff, 2009], for English [Kilgarriff, 1997] served as material.

Thus, it was obtained 15 samples for each language (3 types of words—Swadesh, frequent and random x 5 amounts of the nearest neighbours)—words and corresponding numbers from 0 to 1, denoting % of the stable neighbours. A statistical Mann—Whitney U-test [Mann, Whitney, 1947] was used to evaluate the differences between two independent samples based on the level of any trait measured quantitatively (simple non-parametric criterion).

On each triple of samples (Swadesh, frequent words, random words), a test was conducted with an alternative hypothesis that the values in the second sample were larger. The obtained result for each window of the nearest neighbours is the same:

1. words from Swadesh's list have a higher percentage of stored neighbours than random frequency words from the top 2000;
2. words from Swadesh's list have a higher percentage of saved neighbours than random words of a language;
3. frequency words from the top 2000 have a greater percentage of saved neighbours than just random words of the language[9].

The p-value for all such tests clearly shows that the values in Swadesh's samples are significantly larger than values in frequency word lists; frequency word values are in turn larger than values in random word lists.

(3) *for 100 nearest word neighbours for English:*
$fr = frequent,\ sv = svodesh,\ rn = random$
$rn \leq sw$
$annwhitneyuResult(statistic = 8932.0,\ pvalue = 4.4298335409745345e - 11)$
$fr \leq sw$
$MannwhitneyuResult(statistic = 13363.0,\ pvalue = 0.04262714406973201)$
$rn \leq fr$
$MannwhitneyuResult(statistic = 9962.0,\ pvalue = 3.771709496130687e - 08)$

(4) *for 100 nearest word neighbours for Russian:*
$fr = frequent,\ sv = svodesh,\ rn = random$
$rn \leq sw$
$MannwhitneyuResult(statistic = 8931.0,\ pvalue = 2.4298335409745345e - 11)$
$fr \leq sw$
$MannwhitneyuResult(statistic = 13363.0,\ pvalue = 0.05262714406973201)$
$rn \leq fr$
$MannwhitneyuResult(statistic = 7344.0,\ pvalue = 1.6367105050242702e - 11)$

[9] More complete numbers can be found https://github.com/TatianaShavrina/wordvector_metrics.

3.2. Experiment 2

Further, it was decided to scale up the previous experiment for the entire vocabulary of the existing models and conduct a test on the most stable words model, sorting them all one by one.

The intersection of dictionaries of all models was obtained, then for each word from the list, the number of stable neighbours was calculated- in the window of the 100 nearest neighbours. The list has been sorted by percentage of saved neighbours, remaining the same regardless of model—to measure that the intersection of the list of N nearest neighbours of the word was used on the entire list of models.

Thus, 2 interesting results were obtained at once—at the top of the list, we get the most stable words, which, regardless of the corpus source, keep their neighbours, and at the bottom—the most unstable ones. It is curious that the semantically given top of the list is grouped into distinct semantic groups:

- nouns denoting the personal qualities of a person,

(5) *Russian:*

находчивость_NOUN	*(resourcefulness_NOUN)*	*0.2781 neighbours saved*
радушие_NOUN	*(welcome_NOUN)*	*0.2670*
аккуратность_NOUN	*(accuracy_NOUN)*	*0.2626*
идеализм_NOUN	*(idealism_NOUN)*	*0.2542*

- emotions,

(6) *Russian:*

неприязнь_NOUN	*(hostility_NOUN)*	*0.3059 neighbours saved*
недоверие_NOUN	*(distrust_NOUN)*	*0.2832*
восхищение_NOUN	*(admiration_NOUN)*	*0.2528*
негодование_NOUN	*(resentment_NOUN)*	*0.2473*

- nationalities,

(7) *Russian:*

итальянец_NOUN	*Italian_NOUN*	*0.2558*
ирландец_NOUN	*Irish_NOUN*	*0.2690*
узбек_NOUN	*Uzbek_NOUN*	*0.2389*

- professions,

(8) *Russian:*

скрипач_NOUN	*violinist_NOUN*	*0.2193*
палеонтолог_NOUN	*paleontologist_NOUN*	*0.2179*
филолог_NOUN	*philologist_NOUN*	*0.2391*
географ_NOUN	*geographer_NOUN*	*0.2320*

- toponyms,

(9) *Russian:*

казах_NOUN	*Kazakh_NOUN*	*0.2444*
нижегородский_ADJ	*Nizhny Novgorod_ADJ*	*0.2432*
бразилия_PROPN	*Brazil_PROPN*	*0.2350*
испанский_ADJ	*spanish_ADJ*	*0.2337*

- term adjectives.

(10) *Russian:*

двухлетний_ADJ	*two-year_ADJ*	*0.2428*
четырехмесячный_ADJ	*four month_ADJ*	*0.2278*
трехдневный_ADJ	*three-day_ADJ*	*0.2240*
шестимесячный_ADJ	*six month_ADJ*	*0.2198*

Results are stable for Russian and English (see appendix 1 and appendix 2 correspondingly). Only a few words they are knocked out of a list and can not be assigned to any group: these are 'pregnancy' (0.1386), 'whale' (0.1268), 'intercourse' (0.1226), 'waste' (0.1208) for English, 'неразбериха' ('confusion', 0.2431) 'материализм' ('materialism', 0.2228), 'коррупция' ('corruption', 0.2193) for Russian. There are practically no verbs in the top of the list, for both Russian and English they have too diverse contexts. All the above-mentioned semantic categories were postulated while analyzing the list, the reverse statement that all the words of these categories on average have more stable contexts is not proven because of the difficulty of demarcating these categories.

The most unstable group of words is:

- proper names

(11) *Russian:*

Неклюдов_PROPN	*Neklyudov PROPN*	*0*
Свинцов_PROPN	*Svintsov_PROPN*	*0*
Софронов_PROPN	*Sofronov_PROPN*	*0*
Робсон_PROPN	*Robson_PROPN*	*0*

Having the most inconsistent contexts and low frequency, the proper names—surnames, full names occupy the bottom of the list for both Russian and English.[10]

It is noteworthy that these results partially reproduce the results of clustering in the work [Zobnin, 2017], where groups of proper names, toponyms and other semantic categories are also distinguished.

4. First steps to a linguistic assessment of models

Learning more about the standard properties of a wide list of lexemes in a language, we can more accurately assess both the adequacy of specific models for applied problems and the perspective of their potential improvement.

In the next experiment, we will show how the most popular models for the Russian language retain ontological relations in the vector space. The ontology of RuWordNet [Loukachevitch, Lashevich, 2016], containing more than 300 thousand pairs of words connected by relationships, was taken as a bank of such relations:

POS-synonymy, antonym, cause, domain, entailment, hypernym, hyponym, instance hypernym and instance hyponym, part holonym, part meronym.

---

[10] See full lists at https://github.com/TatianaShavrina/wordvector_metrics.

4 popular word2vec models for Russian—based on News, Aranea, RNC+wiki, Taiga—were studied on the subject of 1) the presence of words in the dictionary, 2) % of the preservation of connections between words—the "presence of a word in the list of N closest neighbours". N is 10, 20, 50, 100. Multi-word expressions are also included in the test—see Table 1.

Table 1: Experimental data examined

child_word	parent_word	relation	has_in_10	has_in_20	has_in_50	has_in_100
рабочий, работник физического труда (worker)	каменщик (mason)	hypernym	FALSE	FALSE	TRUE	TRUE
промышленность (industry)	каменщик (mason)	domain	FALSE	FALSE	FALSE	FALSE

We have 3 values for each word2vec model—"False"—both words presented in a model, but no relation found, 'OOV'—out of vocabulary, one of the words is not presented in a model, 'True'—both words presented in a model, relation established through N nearest words.

The results are surprising in some ways: first, all the metrics turned out to be quite low. Synonymy and antonymy, so beautifully illustrated with examples of original articles, generally stop reproducing for most of the vocabulary. Secondly, the best quality is shown by the model obtained on the largest corpus, Aranea (internet-crawled data), while the model of the Russian National Corpus and Wikipedia shows results below average. The results are also reproduced for the 100 nearest neighbour words (table 2). However, a model trained on the Russian National Corpus and Wikipedia has one of the most comprehensive dictionaries—the number 'not in vocabulary' in it is the smallest in almost all relationships (shown in bold).

Table 2: Remaining % of ontological relations on popular word2vec models, Russian. 100 nearest neighbours

relation	value	taiga	me	news	aranea	mean
antonymy	FALSE	73.05	57.47	75.65	45.56	62.93
antonymy	**TRUE**	25.11	37.77	16.78	**48.70**	32.09
antonymy	OOV	**1.84**	4.76	7.58	5.74	4.98
cause	FALSE	68.44	55.15	78.24	31.23	58.26
cause	**TRUE**	19.44	**41.03**	10.13	15.28	21.47
cause	OOV	12.13	**3.82**	11.63	53.49	20.27
domain	FALSE	96.91	92.73	96.41	90.00	94.01
domain	**TRUE**	1.75	5.80	3.19	**8.13**	4.72
domain	OOV	1.34	1.47	**0.41**	1.87	1.27
entailment	FALSE	91.92	80.13	89.32	64.44	81.45
entailment	**TRUE**	4.46	**17.73**	6.78	12.81	10.45
entailment	OOV	3.62	**2.14**	3.90	22.75	8.10

relation	value	taiga	me	news	aranea	mean
hypemym	FALSE	88.95	82.18	87.64	61.34	80.03
hypemym	**TRUE**	6.81	14.33	7.73	**19.71**	12.15
hypemym	OOV	4.24	**3.49**	4.63	18.95	7.83
hyponym	FALSE	81.57	76.49	79.87	63.20	75.28
hyponym	**TRUE**	7.87	13.40	7.18	**17.75**	11.55
hyponym	OOV	10.57	**10.11**	12.95	19.06	13.17

The lowest quality is shown by popular vector models when conveying relationships like instance hyponymy and domain—it is possible that low quality, among other factors, can be explained by a low frequency of individual occurrences and their absence in the model dictionary. Also, the hypothesis that such relations as hyponymy, hypernymy and domain should be expressed by nearest neighbours can be too simplifiying, as they are hierarchical relations that cannot be extracted from embeddings directly by cosine similarity, unlike pairwise-equivalent synonymy relations.

Relationships of antonymy are fairly well preserved (48% on the best model, Aranea), cause (41% on a best model, Aranea), hypernym, part holonym и part meronym (20% each on Aranea)—but such quality can be considered rather low. Nonetheless, in a similar experiment for the English language [Rogers et al. 2018] skip-gram models show lower quality—synonyms—0.447, antonyms—0.144, hyponyms—0.038, other relations—0.013 of ontological relations.

1) a modern amount of data is still not enough—we need at least an order of magnitude more data to get a large number of contexts for low-frequency words and multi-word expressions, which can be distinguished in a large number in any language;
2) the efficiency of the vectors obtained is far from ideal—between words that are obviously close to the claimed hypothesis: synonyms and antonyms, as well as part-whole and class-subclass relations—the proportion of the saved relations is low.

5. Further work and discussion

Within the framework of the initiated methodology, it is planned to further study the distributional lexis behaviour, and based on the results obtained, it is planned to develop metrics that allow obtaining a more complete interpretation of vector models.

The author would like to start a discussion on whether vector models can be used as a tool for a full-fledged linguistic lexical study on big corpora: potentially, such areas of study could be:

- assessment of corporal context biases, corpus thematic focus
- assessment of the sufficiency of the presented contexts of basic vocabulary in the corpus
- search for the most universal vocabulary groups that preserve the structure of relations among themselves regardless of the corpus and model
- the formation of a clearer picture of the set of mandatory properties that characterize a representative corpus of a language.

6. Conclusion

In this paper, as a result of experiments conducted on popular word2vec models for Russian and English, it was shown that the most stable lexical groups, having most uniform contexts, independent from the corpus, are:

- adjectives denoting the personal qualities of a person,
- nationalities,
- professions,
- toponyms,
- term adjectives.

At the same time, the most unstable group are proper names—as the rarest and context-dependent.

It has been established that words from Swadesh list (for English and Russian) are more resistant to a change of model and retain their closest neighbours regardless of the model much more often than words from the frequency vocabulary, as well as more often than random words.

For the Russian language, an experiment was conducted to assess the residual number of semantic and ontological links between known pairs of words and the quality of models was estimated on the basis of this number of relations remaining in the model.

All the data and code for this paper are available on github[11]—we welcome other authors to contribute word2vec metrics and evaluate their models.

7. Acknowledgement

The author is sincerely grateful to Olga Lyashevskaya and Serge Sharoff who prompted the author to think about the need for a different methodology for evaluating vector models, to Andrei Kutuzov for providing additional information about RusVectores models, and to Natalia Lukashevich for providing materials from the RuWordNet project.

References

1. *Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean* (2013). Efficient estimation of word representations in vector space. CoRR, abs/1301.3781.
2. *Jeffrey Pennington, Richard Socher, and Christopher D. Manning* (2014). GloVe: Global Vectors for Word Representation.
3. *P. Bojanowski*, E. Grave*, A. Joulin, T. Mikolov* (2017) Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics, volume 5, 2017, issn 2307–387X, pp. 135–146
4. *Ferdinand de Saussure* (1916) Cours de linguistique générale, ed. C. Bally and A. Sechehaye, with the collaboration of A. Riedlinger, Lausanne and Paris: Payot; trans. W. Baskin, Course in General Linguistics, Glasgow: Fontana/Collins, 1977.

11 https://github.com/TatianaShavrina/wordvector_metrics.

5. *Kutuzov A., Kuzmenko E.* (2017) WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In: Ignatov D. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science, vol 661. Springer, Cham
6. *Kutuzov, A., Andreev, I.* (2015) Texts in, meaning out: neural language models in semantic similarity task for Russian. In: Proceedings of the Dialog Conference, Moscow, RGGU
7. *Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.* (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv e-prints
8. *Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer* (2018). Deep contextualized word representations. In NAACL
9. *Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever* (2018). Improving language understanding with unsupervised learning. Technical report, OpenAI.
10. *Swadesh, Morris.* (1950). “Salish Internal Relationships.” International Journal of American Linguistics, Vol. 16, 157–167.
11. *О. Н. Ляшевская, С. А. Шаров,* (2009). Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник.
12. *Loukachevitch N., Lashevich G.* (2016) Multiword expressions in Russian Thesauri RuThes and RuWordNet. Proceedings of the AINL FRUCT 2016, pp. 66–71.
13. *Kilgarriff, A.* (1997) Putting Frequencies in the Dictionary. International Journal of Lexicography 10 (2). Pp 135–155.
14. *Mann, Henry B.; Whitney, Donald R.* (1947). “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other”. Annals of Mathematical Statistics. 18 (1): 50–60. doi:10.1214/aoms/1177730491. MR 0022058. Zbl 0041.26103.
15. *Felix Hill, Roi Reichart, and Anna Korhonen* (2016). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. Computational Linguistics
16. *Rogers, Anna, Shashwath Hosur Ananthakrishna, and Anna Rumshisky* (2018). What’s in your embedding, and how it predicts task performance. In Proceedings of the 27th International Conference on Computational Linguistics, pages 2690–2703.
17. *Yulia Tsvetkov, Manaal Faruqui, and Chris Dyer* (2016). Correlation-based intrinsic evaluation of word vector representations. CoRR abs/1606.06710.
18. *Ivan Vulic, Daniela Gerz, Douwe Kiela, Felix Hill, and Anna Korhonen* (2016). Hyperlex: A large-scale evaluation of graded lexical entailment. CoRR,abs/1608.02117.
19. *A. Zobnin* (2017) “Rotations and interpretability of word embeddings: the case of the russian language,” arXiv preprint arXiv:1707.04662.

Appendix

Appendix 1 and 2: Top 100 most stable words for English and Russian

N top	intersection	word English	intersection	word Russian
1	0.2658959538	six-month_ADJ	0.3058823529	неприязнь_NOUN
2	0.2388888889	three-week_ADJ	0.2832369942	недоверие_NOUN
3	0.2131147541	two-week_ADJ	0.2781065089	находчивость_NOUN
4	0.1935483871	eight-year_ADJ	0.269005848	ирландец_NOUN
5	0.1917098446	six-week_ADJ	0.2670454545	радушие_NOUN
6	0.1917098446	two-year_ADJ	0.2625698324	аккуратность_NOUN
7	0.1904761905	four-month_ADJ	0.2558139535	итальянец_NOUN
8	0.1808510638	three-month_ADJ	0.2542372881	идеализм_NOUN
9	0.1804123711	extremely_ADV	0.2528089888	изобретательность_NOUN
10	0.175879397	Uganda_PROPN	0.2528089888	восхищение_NOUN
11	0.1693121693	unease_NOUN	0.25	самоотверженность_NOUN
12	0.1691542289	Malawi_PROPN	0.2472527473	негодование_NOUN
13	0.1675126904	four-week_ADJ	0.2471264368	добросовестность_NOUN
14	0.16	Tanzania_PROPN	0.2458100559	расторопность_NOUN
15	0.1592039801	seven-day_ADJ	0.2445652174	невероятный_ADJ
16	0.158974359	resentment_NOUN	0.2444444444	казах_NOUN
17	0.1565656566	disappointment_NOUN	0.2432432432	нижегородский_ADJ
18	0.1534653465	Botswana_PROPN	0.2430939227	неразбериха_NOUN
19	0.1507537688	astonishing_ADJ	0.2427745665	двухлетний_ADJ
20	0.1477832512	immense_ADJ	0.2391304348	филолог_NOUN
21	0.1469194313	Zambia_PROPN	0.2388888889	узбек_NOUN
22	0.1464646465	Guyana_PROPN	0.2369942197	десятилетний_ADJ
23	0.1464646465	homosexual_ADJ	0.2362637363	недовольство_NOUN
24	0.1463414634	incompetence_NOUN	0.2349726776	бразилия_PROPN
25	0.1435897436	violin_NOUN	0.2349726776	сметка_NOUN
26	0.1428571429	Mozambique_PROPN	0.2346368715	неодобрение_NOUN
27	0.1428571429	five-day_ADJ	0.2336956522	испанский_ADJ
28	0.1407035176	ten-year_ADJ	0.2329545455	смекалка_NOUN
29	0.14	inaccurate_ADJ	0.2320441989	географ_NOUN
30	0.14	three-year_ADJ	0.2316384181	настойчивость_NOUN
31	0.1386138614	pregnancy_NOUN	0.2316384181	грузин_NOUN
32	0.1379310345	five-week_ADJ	0.2315789474	румыния_PROPN
33	0.1359223301	incredible_ADJ	0.2311827957	вологодский_ADJ
34	0.1359223301	Grenada_PROPN	0.2307692308	оплошность_NOUN
35	0.1359223301	tedious_ADJ	0.2295081967	ирландский_ADJ
36	0.1355140187	amazing_ADJ	0.2287234043	омский_ADJ
37	0.1355140187	Kenya_PROPN	0.2287234043	финн_NOUN
38	0.1346153846	enormous_ADJ	0.2277777778	дружелюбие_NOUN
39	0.1346153846	shocked_ADJ	0.2277777778	четырехмесячный_ADJ
40	0.1343283582	disquiet_NOUN	0.2272727273	лицемерие_NOUN
41	0.1339712919	Lesotho_PROPN	0.227027027	саратовский_ADJ
42	0.1333333333	Sierra::Leone_PROPN	0.226519337	азербайджанец_NOUN
43	0.1320754717	dismay_NOUN	0.226519337	католик_NOUN
44	0.1320754717	Zimbabwe_PROPN	0.2263157895	венгрия_PROPN
45	0.1317073171	greatly_ADV	0.2252747253	спокойствие_NOUN
46	0.1306532663	appalling_ADJ	0.2252747253	пунктуальность_NOUN
47	0.1280788177	alarmed_ADJ	0.2247191011	бесполезность_NOUN

N top	intersection	word English	intersection	word Russian
48	0.1279620853	remarkable_ADJ	0.2247191011	некомпетентность_NOUN
49	0.1267605634	whale_NOUN	0.2245989305	ростовский_ADJ
50	0.1261682243	incredibly_ADV	0.2240437158	трехдневный_ADJ
51	0.125	Antigua_PROPN	0.2240437158	деликатность_NOUN
52	0.125	teenager_NOUN	0.222826087	чудовищный_ADJ
53	0.1237623762	honesty_NOUN	0.222826087	материализм_NOUN
54	0.1231527094	Nigeria_PROPN	0.222826087	индус_NOUN
55	0.1231527094	ankle_NOUN	0.2216216216	трехнедельный_ADJ
56	0.1227272727	biologist_NOUN	0.2215909091	американец_NOUN
57	0.1226415094	intercourse_NOUN	0.2204301075	недоумение_NOUN
58	0.1225490196	Dominica_PROPN	0.2204301075	австралия_PROPN
59	0.1218274112	frustration_NOUN	0.2197802198	шестимесячный_ADJ
60	0.1213592233	underwear_NOUN	0.2192513369	скрипач_NOUN
61	0.1209302326	Barbados_PROPN	0.2192513369	коррупция_NOUN
62	0.1207729469	waste_NOUN	0.217877095	палеонтолог_NOUN
63	0.1207729469	trumpet_NOUN	0.2173913043	неясность_NOUN
64	0.1201923077	generosity_NOUN	0.2173913043	неимоверный_ADJ
65	0.1196172249	clarinet_NOUN	0.2162162162	пакистан_PROPN
66	0.119266055	conspiracy_NOUN	0.2159090909	биолог_NOUN
67	0.119266055	whisky_NOUN	0.2157894737	недельный_ADJ
68	0.1188118812	cello_NOUN	0.2154696133	антисемитский_ADJ
69	0.1184834123	courage_NOUN	0.2150537634	венгерский_ADJ
70	0.117370892	sex_NOUN	0.2142857143	предусмотрительность_NOUN
71	0.117370892	surgeon_NOUN	0.2131147541	презрение_NOUN
72	0.1170731707	pear_NOUN	0.2131147541	усидчивость_NOUN
73	0.1165048544	nine-year_ADJ	0.2131147541	дотошный_ADJ
74	0.1165048544	ten-day_ADJ	0.2124352332	возмущение_NOUN
75	0.1165048544	Nairobi_PROPN	0.2116402116	таджик_NOUN
76	0.1162790698	flute_NOUN	0.2116402116	ирландия_PROPN
77	0.1162790698	headache_NOUN	0.2111111111	этнограф_NOUN
78	0.1157407407	uncle_NOUN	0.2111111111	сноровка_NOUN
79	0.1153846154	craftsman_NOUN	0.2108108108	геолог_NOUN
80	0.1148325359	sadness_NOUN	0.2105263158	армянин_NOUN
81	0.1148325359	weather_NOUN	0.2099447514	выразительность_NOUN
82	0.1142857143	t-shirt_NOUN	0.2099447514	ангола_PROPN
83	0.1141552511	marvellous_ADJ	0.2099447514	православие_NOUN
84	0.1141552511	frustrating_ADJ	0.2099447514	плечистый_ADJ
85	0.1136363636	biology_NOUN	0.2096774194	ярославский_ADJ
86	0.1132075472	despair_NOUN	0.2096774194	тщательность_NOUN
87	0.1132075472	consternation_NOUN	0.2096774194	симпатия_NOUN
88	0.112745098	concerto_NOUN	0.2096774194	грузинский_ADJ
89	0.1126760563	sexual_ADJ	0.2094240838	пермский_ADJ
90	0.1126760563	perseverance_NOUN	0.2087912088	дагестанский_ADJ
91	0.1126760563	husband_NOUN	0.2087912088	голландец_NOUN
92	0.1126760563	inventiveness_NOUN	0.2087912088	беспокойство_NOUN
93	0.1126760563	arduous_ADJ	0.2085561497	зависть_NOUN
94	0.1126760563	false_ADJ	0.2085561497	злорадство_NOUN
95	0.1126760563	homosexuality_NOUN	0.2085561497	невероятно_ADV
96	0.1121495327	tuna_NOUN	0.2085561497	томский_ADJ
97	0.1121495327	frequently_ADV	0.2078651685	никчемность_NOUN
98	0.1121495327	rivalry_NOUN	0.2076502732	коренастый_ADJ
99	0.1116504854	shirt_NOUN	0.2076502732	воронежский_ADJ