Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2016"

Moscow, June 1–4, 2016

THE IMPACT OF DIFFERENT DATA SOURCES ON FINDING AND RANKING SYNONYMS FOR A LARGE-SCALE VOCABULARY

Antonova A. (antonova@yandex-team.ru),
Kobernik T. (kobernik@yandex-team.ru),
Misyurev A. (misyurev@yandex-team.ru)

Yandex, Moscow, Russia

In this paper we compare different models for measuring synonymy. We consider methods based on monolingual text corpora and parallel texts. We experiment with the features based on context similarity, translation similarity, and similarity of neighbors in the parse trees. We provide an analysis of strong and weak points of different approaches and show that their combination can improve the results. The considered methods can handle large-scale vocabularies and be useful for automatic construction of human-oriented synonym dictionaries.

Keywords: synonyms, semantic similarity, synonym dictionary extraction, vector models, translation models

1. Introduction

In this paper we compare different models for measuring synonymy. Such methods can be useful for automatic construction of synonym dictionaries (see Fig. 1).

We consider methods based either on plain monolingual texts, or on parallel texts. Such corpora can be gathered from the Web and updated regularly with the growing number of documents. The automatically constructed synonym dictionaries can have the following advantages: coverage, variety, sensitivity to real occurrence in texts, and recent language changes. For this reason, it is interesting to compare automatically extracted synonyms and the synonyms from a human-built dictionary.

Many existing methods of automatic extraction of synonyms are based on the similarity of contexts in monolingual texts (see 2.2). However, many of the reported experiments are of a preliminary nature. The dependence between the quality of the results and the amount of data and the word frequencies is not quite understood. Context-based methods have some limitations, since the context similarity does not always directly correspond to synonymy. Problems can arise in the case of polysemous words. It is also difficult to collect reliable contextual information for rare words.

ум сущ

разум · рассудок · интеллект · смысл · здравый смысл · гений · разумение · толк
голова · мозг · память · головной мозг
догадка · мысль · мышление · взгляд · мнение · намерение · соображение
сметка · догадливость · остроумие · смекалка · сообразительность · смышленость
рассудительность · трезвость
мыслительные способности · умственные способности
интеллектуальность · интеллигентность

Fig. 1. Example of synonym dictionary entry for word «ум» ("mind, intellingence")

Another useful source of information about synonymy can be found in parallel corpora. Statistical translation models (phrase-tables) can serve as a source of synonym candidate pairs. One can assume that if two words have common translations in another language, they can be synonyms. Translation frequencies in the phrase-tables can be used to estimate the distance between synonym candidates.

It seems that having the information about the syntactic relations between words can also be useful [Dekang Lin 1997]. One could use the same context-based methods, but replace context words with syntactically related words, i.e. words adjacent in the parse tree. Such data is likely to contain less noise.

The aim of the paper is to study the impact of different data sources (namely, vector models, translation tables, syntax) on the quality of finding and ranking synonyms. We are interested in obtaining large and accurate human-oriented dictionaries. We are also interested in studying the dependence of the results on the size of the corpora, the size of vocabulary and word frequencies. We report on the experiments with combinations of different data sources for Russian.

The notion of synonymy is rather vague and subjective, which makes it difficult to find a reliable formal metric. The evaluation metrics based on assessments of human experts can be hard to reproduce. It is not clear how such metrics take into account the specifics of possible areas of practical application. For the evaluation of automatically extracted synonyms, we compare them to the synonyms from human-built dictionaries.

We restrict the pairs of words that can be considered candidates for synonyms, to the words that have at least one common translation in an SMT phrase-table. It turns out that part of manual synonym pairs cannot be found in this way [see Section 4.3]. On the other hand, the automatically extracted synonyms can be more relevant and up-to-date, since they reflect the word usage in real texts.

We regard as reference all synonym candidate pairs that intersect with human-built dictionary. The rest can contain both correct and not correct synonym pairs. A significant portion of automatically generated correct synonym pairs may not appear in the reference. In this approach, the task of finding the right synonym pairs is similar to the problem of ranking. Different models of similarity estimation produce the ranked lists of candidate pairs, which then can be compared w.r.t. the ranks of reference pairs. We believe that this experimental setup may be useful for the evaluation and analysis of different methods. On the one hand, we rely on publicly available dictionary and do not need any further human assessment. On the other hand, the task has clear practical value.

The rest of the paper is organized as follows. In Section 2 we outline the related work. In Section 3 we describe the models of similarity estimation, and the features that we use. In Section 4 the experimental setup is described. The results of the experiments are reported in Section 5. We conclude and discuss the applicability of overall approach to building large-scale synonym dictionaries in Section 6.

2. Related work

The task of synonym extraction is closely related to the more general problem of measuring semantic similarity of words and phrases. The existing approaches can be roughly subdivided into three types according to the main source of information about semantic similarity.

2.1. Knowledge-based approaches

Knowledge-based methods try to make use of existing lexical resources, such as thesauri or knowledge graphs that represent a kind of a semantic network. The similarity of two words can be estimated, taking into account the distance between them in this network. Many approaches[Richardson 1994, Postma 2014] are based on Wordnet[Miller 1995, Fenenbaum 1998], a manually created lexical resource for English, but there also exist attempts to automatically induce semantic networks, e.g. from Wikipedia. However, such methods are left beyond the scope of this paper, since they require large-scale resources to be available for a particular language.

2.2. Monolingual context-based approaches

There exist different methods of measuring semantic similarity between two words based on the lexical coocurrence in large monolingual corpora. A wide variety of measures were proposed [see Baroni 2014 for systematic comparison] from simple scalar product of coocurrence frequency vectors, or Kullback-Leibler distance between the context distributions to more complex methods, that overcome the data sparsity problem, such as Latent Semantic Analysis (LSA) [Landauer and Dumais 1997], Distributional Memory [Baroni 2009] and neural network language models [Mikolov 2013, Pennington 2014].

The application of different methods to Russian is discussed in detail in the materials of Russe-2015 [Panchenko et al. 2015] contest. Different approaches are presented, including distributional, and neural network-based models, trained on a wide variety of monolingual corpora, as well as knowledge-based approaches.

2.3. Translation-based approaches

There exist methods for the extraction of synonym candidate pairs from bilingual parallel corpora. They use the alignment techniques from phrase-based statistical machine translation, and identify candidate synonyms using a phrase in another language as a pivot. [Dolan 2004, Bannard 2005, Barzilay 2001, Zhao 2008, Bansal 2012].

3. Models for similarity estimation

3.1. Translation model and extraction of synonym candidate pairs

A translation model $TM = \left(en_i, ru_j, count\right)$ is the set of translation equivalents with their respective counts. The translation equivalents are extracted from a parallel corpus with the help of SMT techniques and tools [Koehn 2003]. We also preprocess parallel sentences by a morpho-syntactic analyzer [Antonova, Misyurev 2012]. This allows us to sum over the counts the translations with the same lemmas, selecting only one pair in the translation model as described in [Antonova, Misyurev 2014].

The set of Russian synonym candidate pairs can be defined as

$$Cand(t) = \{(ru_a, ru_b) \mid \exists en_i, c_a, c_b, (en_i, ru_a, c_a) \in$$

$$\in TM, (en_i, ru_b, c_b) \in TM, c_a \geq t, c_b \geq t\} \quad (1)$$

Eq. 1 describes the set of pairs of Russian words for which there exists at least one common translation with a joint count above a given threshold t.

3.2. Translation similarity score

The similarity estimate for a pair of synonym candidate pair is calculated by Eq. 2.

$$TranslationSimilarity(ru_a, ru_b) = TranslationSimilarity(ru_b, ru_a) =$$

$$= Pr(ru_a|ru_b)Pr(ru_b|ru_a) \quad (2)$$

$$Pr(ru_a|ru_b) = \sum_i Pr(ru_a|en_i)Pr(en_i|ru_b) \quad (3)$$

$$Pr(ru_b|ru_a) = \sum_i Pr(ru_b|en_i)Pr(en_i|ru_a) \quad (4)$$

$$Pr(en_i|ru_a) = \frac{Count(en_i, ru_a)}{\sum_{en_i} Count(en_i, ru_a)} \quad (5)$$

For all common English translations en_i we calculate the probabilities of translating ru_a to ru_b through en_i, and then marginalize over en_i to get $Pr(ru_a|ru_b)$. To get a single symmetrical similarity estimate we get a multiplication of $Pr(ru_a|ru_b)$ and $Pr(ru_b|ru_a)$.

Typical mistakes observed when ranking synonym candidate pairs by translation similarity score are the following:

- Mistakes of automatic word alignment may produce incorrect translations pairs. Particularly, wrong pieces of multiword expressions can be found in the phrase table with high counts.
- Mistakes introduced by the word sense ambiguity. For example, unrelated Russian words «бежать» (“move quickly”) and «запускать» (“launch”) can be translated by a polysemous English word “run”. Such polysemous common translations can connect words that are not synonyms.

These two types of mistakes are specific for the phrase-table and one can expect that a combined approach taking monolingual contexts into account can improve the ranking. It is important to use a lemmatized phrase-table, otherwise the synonym candidate pairs can contain many word forms of the same lemma and the counts can be much sparser.

3.3. Similarity scores by vector models

For measuring similarity in monolingual contexts we train vector models with the help of two popular tools *word2vec* and *glove*. They represent each word as a vector in a low-dimensional space. The similarity score between two words is given by the scalar product of the corresponding vectors. Though vector models are widely popular and effective, this approach still has some weaknesses:

- It cannot divide separate meanings of polysemous words, since each word has only one corresponding vector.
- Vectors for rare words can be unreliable, since they occur in a small number of contexts in the corpus.
- It is appropriate for single words, but requires special efforts to handle multiword expressions.

3.4. Vector models with syntactically related words

We checked the possibility of using syntactic relations to construct vector models of words. As is known, such models are based on the co-occurrence statistics of the corpus. Instead of collecting all words within a predefined window, we collected syntactic contexts, i.e. the words that are adjacent in the parse tree. Such data is likely to contain less noise. We trained a *glove* vector model on the set of syntactic contexts. In the rest of the paper this model is referred to as *GloveSynt*.

3.5. Model combination

It seems that context similarity and translation similarity suffer from different types of problems and that their combination can improve the results. Moreover, taking word frequencies into account can possibly lead to further improvement. We combine the similarity scores by different models and word frequencies in a log-linear model, and train the weights with logistic regression.

3.6. Quality metrics

The notion of synonymy is rather vague and subjective, which makes it difficult to find a reliable formal metric. To assess the quality of a ranked list of automatically extracted synonyms, we look at the ranks of the gold synonyms from a human-built dictionary. We evaluate the importance of different features w.r.t. the ranking that they produce on the list of all candidate pairs. To measure the ranking quality we use the following metrics: average precision (*AveP*), average rank(*AveRank*), median of ranks(*Median*). The average precision is defined as follows:

$$AveP = \frac{\sum_{r=1}^{n} P(r) \times rel(r)}{\sum_{r=1}^{n} rel(r)} \quad (6)$$

where r is the rank in the sequence of candidates, $P(r)$ is the precision of top r candidates, $rel(r)$ is an indicator function equaling 1 if r-th pair is relevant, n is the total number of candidates.

4. Experiment

4.1. Reference synonym pairs

We downloaded Russian synonym dictionary from Wiktionary.org, taking only semantic relation of type "Synonym". The initial 58,715 synonym pairs were lowercased, and symmetrized[1] (105,142 pairs after symmetrization). Considering only single-word pairs for the described experiment we got a set of reference pairs, Gold = {(query, synonym)}, which contained 99,394 single-word synonym pairs for 42,509 distinct queries.

Another dataset *GoldAbr* consisted of Abramov's dictionary of Russian synonyms and similar words, whose first edition was in 1915. After symmetrization it contained 34,930 pairs for 12,527 distinct queries. The intersection of Wiktionary and Abramov dictionary is small: 6,616 pairs.

---

[1] Though some synonym pairs (a, b) may be asymmetrical, most added pairs (b, a) are also relevant.

4.2. Synonym candidate pairs

We built a lemmatized phrase-table with maximum phrase length of 3 words on an English-to-Russian corpus drawn from the Web. The minimal joint translation count is 2. The sum of all joint counts is about 2.91 billion. For the simplicity of the experiment, we restricted the Russian side only to single words that had been recognized as correct Russian lemmas by an in-house morphological dictionary.

We generated a set of synonym candidate pairs with the help of the phrase-table, as described in 3.1. The set of positive examples consisted of the intersection of reference and candidate pairs $Pos = Gold \cap Cand$. The set of negative examples consisted of those pairs, whose query word occurred in *Gold* and had at least one positive example in *Pos*.

All query words were randomly divided into two parts. Then all positive and negative examples were place into one of two sets according to their query word:

- training set: 26,281 positive, 2,657,507 negative examples.
- test set: 26,461 positive, 2,614,819 negative examples.

Fig. 2 represents the dependence between the number of candidate synonyms and the query frequency.

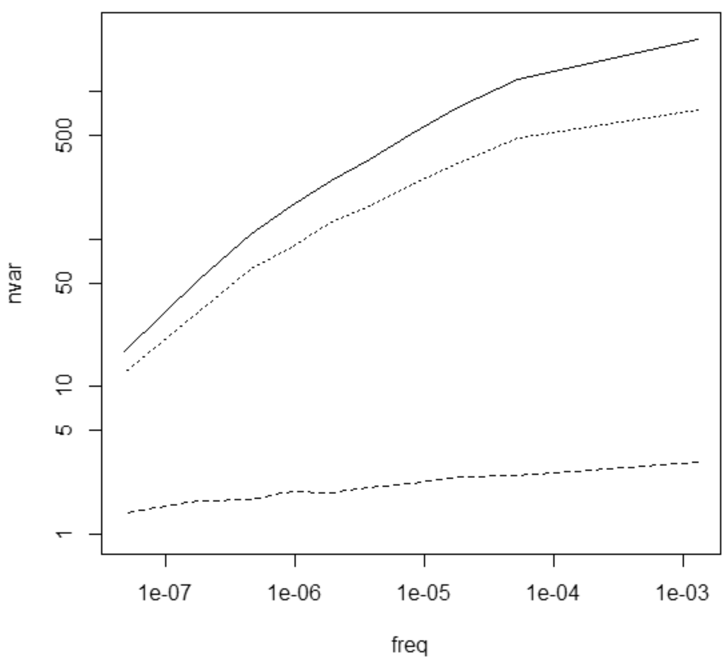

Fig. 2. Dependence between the number of candidate synonyms and the query frequency. Bottom line — reference synonyms, upper line — all candidates by translation model, the middle line — number of candidates per one reference pair

Frequent words typically have more synonyms in human-built dictionaries, and more synonym candidate pairs. The absolute number of candidate pairs is very big (logarithmic scale). On the one hand we have more data for more frequent words. On the other hand the classification task is harder because they have more synonym candidates.

Concerning the task of synonym evaluation, there exists a problem of reference sparsity. It means that the candidate list often contains many relevant synonym pairs that are missing in the reference (see Table 1). For that reason, the use of standard metrics such as recall/precision may be unconvincing.

Table 1. Top-34 synonym candidates for word «проворный» (“agile”), ranked by descending translation similarity. Reference synonyms are bold (Wiktionary) and underlined (Abramov)

верткий	(~nimble)	$1.8e^{-3}$	**бойкий**	(~spirited)	$7.6e^{-6}$
ловкий	(~agile)	$9.4e^{-4}$	незамедлительный	(~immediate)	$7.2e^{-6}$
поворотливый	(~agile)	$8.7e^{-4}$	сноровистый	(~nimble)	$5.2e^{-6}$
шустрый	(~nimble)	$6.6e^{-4}$	подвижной	(~mobile)	$5.0e^{-6}$
юркий	(~nimble)	$2.7e^{-4}$	подсказывающий	(~prompting)	$3.4e^{-6}$
прыткий	(~nimble)	$1.7e^{-4}$	динамичный	(~dynamic)	$2.7e^{-6}$
маневренный	(~maneuvering)	$9.6e^{-5}$	оживленный	(~brisk)	$2.6e^{-6}$
быстрый	(~fast)	$5.4e^{-5}$	своевременный	(~timely)	$2.3e^{-6}$
быстро	(~quickly)	$5.3e^{-5}$	находчивый	(~resourceful)	$1.7e^{-6}$
расторопный	(~agile)	$5.2e^{-5}$	изворотливый	(~quirky)	$1.6e^{-6}$
гибкий	(~flexible)	$4.7e^{-5}$	безотлагательный	(~urgency)	$1.4e^{-6}$
вертлявый	(~fidgety)	$3.6e^{-5}$	пробужденный	(~awakened)	$8.2e^{-7}$
подвижный	(~mobile)	$3.1e^{-5}$	активный	(~agile)	$8.1e^{-7}$
оперативный	(~operational)	$1.8e^{-5}$	скорый	(~fast)	$8.0e^{-7}$
стремительный	(~rapid)	$1.7e^{-5}$	сообразительный	(~witted)	$8.0e^{-7}$
скорейший	(~early)	$1.4e^{-5}$	**резвый**	(~spirited)	$6.3e^{-7}$
подсказанный	(~prompted)	$8.8e^{-6}$	бодрый	(~brisk)	$6.2e^{-7}$

Fig. 3 illustrates the problem of measuring the ranking quality with the average precision when the reference is sparse. One can see that although some reference pair are ranked high according to our similarity models, a considerable amount of reference pairs belong to the low-precision area. For that reason we used additional metrics, namely, average rank(*AveRank*), median of ranks(*Median*).

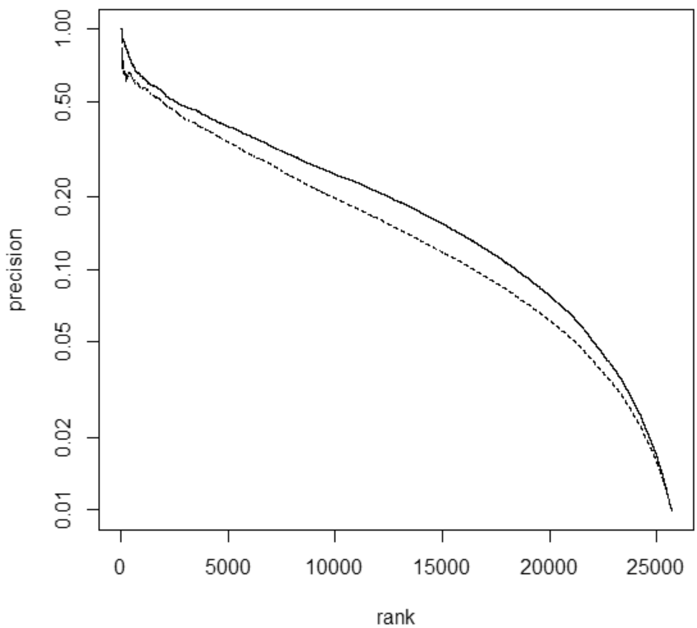

Fig. 3. Precision vs rank. Solid line — translation similarity, dashed line — word2vec

4.3. Missing synonym pairs

Only 58.9% of initial reference pairs were found among candidates generated by the phrase-table. Among those missing pairs, 61.3% are the pairs in which one or both words had not occurred in the phrase-table at all, 38.7% are the pairs in which both words occurred in the phrase-table, but had no common translations.

We manually annotated 100 random missing pairs for Wiktionary (see Table 2). Only 36% of them were actually good synonyms, though they also included words with low frequencies.

Table 2. Human annotation results for 100 random reference pairs without common translations

%	Human judgement	Example
36	Good candidates	ряженка — варенец (~milk product), пацанва — детвора (~children)
25	Both words are rare or unknown	тонемика — тонология (~tonology?)
21	One word is rare or unknown	гуртоправ — гуртовщик (~drover), скворец — кокако (~starling)
6	One word is slang or obscene	записка — малявка (~note)
5	Words are not synonyms	важный (~important: adjective) — царственно (~kingly: adverb)
4	One word is not Russian	галоген — галоїд (~halogen)
3	Obsolete meaning	сплетник (~gossip) — трубач (~trumpeter), управлять (~to control) — рядить (~to dress?)

4.4. Features for model combination

We calculated the following features for each synonym candidate pair from test set and training set.

1. Logarithm of translation similarity score (see Eq.2).
2. Scalar product of vectors by the word2vec model. The model was trained with standard parameters, the vocabulary consisted of words occurring at least 10 times in the corpus. The corpus consisted of 200 mln sentences, disambiguated and lemmatized with a morpho-syntactic analyzer [Antonova, Misyurev 2012].
3. Scalar product of vectors by the glove model. The training setting was the same as previous.
4. Logarithms of the frequencies of the two synonyms in the monolingual corpus.
5. Scalar product of vectors by the glove model trained on syntactic contexts. The contexts include lemmas that are adjacent to the given word in the parse trees.

We combined the combinations of the above features in a log-linear model, and trained the weights with logistic regression.

5. Experiment results

We report the ranking quality given by different models and their combinations. A summary of these can be seen in Table 3.

Table 3. Ranking results for different feature combinations. The metrics are average precision (*AveP*), average rank (*AveRank*) and median of ranks (*Median*)

Feature combination	AveP	AveRank	Median
Wiktionary dataset			
Word2Vec	0.165	407,228	132,683
TranslationSimilarity	0.237	222,513	66,802
TranslationSimilarity + Frequencies	0.247	212,819	65,304
Word2Vec + TranslationSimilarity + Frequencies	**0.303**	**181,381**	**50,232**
Glove	0.117	560,219	219,061
Glove + TranslationSimilarity + Frequencies	0.299	182,327	50,338
GloveSynt	0.058	803,457	467,868
GloveSynt + TranslationSimilarity	0.274	204,993	57,393
GloveSynt + TranslationSimilarity + Frequencies	0.291	191,225	54,492
Abramov dataset			
Word2Vec	0.025	516,313	244,381
Word2Vec + TranslationSimilarity + Frequencies	0.068	250,096	79,268
Glove	0.031	506,889	220,102
Glove + TranslationSimilarity + Frequencies	**0.075**	**238,683**	**73,224**
TranslationSimilarity	0.049	272,115	99,969

The vector model with syntactic contexts (*GloveSynt*) yields an improvement in combination with translation similarity model. However, the vector models trained on simple lemmatized text yield better results. Besides, using syntactic contexts requires parsing the corpus, which makes the experiments more complex, time-consuming and difficult to reproduce.

The classification results with the single *glove* model turned to be lower than those of *word2vec* model for Wiktionary dataset, but higher for Abramov dataset. It is interesting that in combination with the translation similarity model, they are almost equal in quality. The advantage of *glove* tool is that it allows us to parallelize the context extraction, e. g. with map-reduce operations.

It seems that Abramov dictionary contains less straightforward synonyms and more distant synonyms than Wiktionary. The absolute values of average precision for Abramov dataset is much lower than that for Wiktionary dataset.

Fig. 4 demonstrates the advantages of combination of different models for classification and ranking. Though they correlate on many examples, using two dimensions makes it possible to classify correctly some uncertain points.

Fig. 5 demonstrates the top-ranked pairs by different models depending on the word frequencies. One can see that the vector models tend to rank higher frequent words, while the translation similarity model ranks better rare words, but tends to prioritize words with similar frequency.

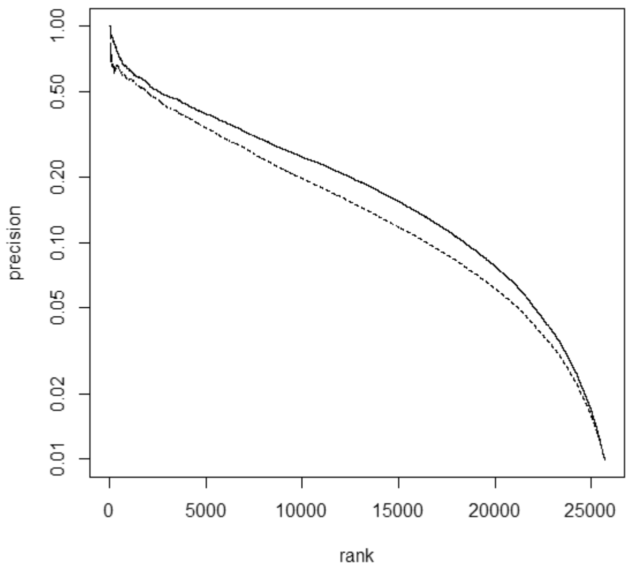

Fig. 4. Joint distribution of positive (○) and negative (×) examples w.r.t. TranslationSimilarity score and word2vec distance (Wiktionary dataset). Pt—is translation similarity score

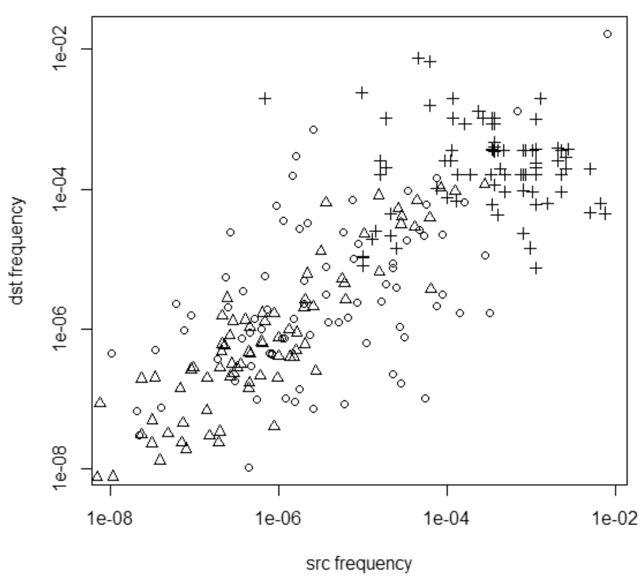

Fig. 5. Distribution of candidate pairs (src, dst) w.r.t. the word frequencies in Wiktionary dataset (○ — random reference pairs, Δ — top by translation similarity, × — top by vector models)

Conclusion

We compare the ranking of synonym candidate pairs given by different models. We consider models based on syntactic relations, monolingual contexts, and models based on parallel texts.

The set of synonym candidate pairs is generated with the help of the phrase-table, which is extracted from parallel texts with SMT methods. The translation similarity model based on the phrase-table statistics also proves to be useful for ranking candidates. It can handle rare and polysemous words.

We show that the precision of different models depends on word frequencies. Our experiments demonstrate that the combination of monolingual vector models and translation similarity model improves the ranking results, as well as taking word frequencies into account. The general approach has a practical value, since it can handle large-scale vocabularies and be useful for automatic construction of synonym dictionaries.

References

1. *Alexandra Antonova and Alexey Misyurev.* (2012), Russian dependency parser SyntAutom at the Dialogue-2012 parser evaluation task. Proceedings of the Dialogue-2012 International Conference.
2. *Alexandra Antonova and Alexey Misyurev.* (2014), Automatic creation of human-oriented translation dictionaries. Proceedings of the Dialogue-2014 International Conference.
3. *Bannard C., Callison-Burch C.* (2005), Paraphrasing with bilingual parallel corpora, Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, p. 597–604, June 25–30, 2005, Ann Arbor, Michigan.
4. *Bansal M., Denero J., Dekang Lin.* (2012), Unsupervised translation sense clustering, Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, June 03–08, Montreal, Canada.
5. *Baroni M., Dinu G., Kruszewski G.* (2014), Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Vol. 1), Baltimore, MD, USA, pp. 238–247.
6. *Baroni M., Lenci A.* (2009), One distributional memory, many semantic spaces. Proceedings of the EACL GEMS Workshop. Athens, Greece, pp. 1–8.
7. *Barzilay R., McKeown K. R.* (2001), Extracting paraphrases from a parallel corpus. ACL'01 Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, pp. 50–57
8. *Dekang Lin.* (1997), Using syntactic dependency as local context to resolve word sense ambiguity. In In Proceedings of ACL-97, page 64–71.
9. *Dolan W., Quirk C., Brockett C.* (2004), Unsupervised Construction of Large Paraphrase Corpora: Exploiting Massively Parallel News Sources, International Conference on Computational Linguistics, August 2004.

10. *Fellbaum C., ed.* (1998), WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA.
11. *Koehn P., Och F. J., Marcu D.* (2003), Statistical phrase-based translation. In Proceedings of NAACL.
12. *Landauer T. K., Dumais S. T.* (1997), A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge, Psychological Review, Vol. 104(2), pp. 211–240.
13. *Mikolov T., Chen K., Corrado G., Dean J.* (2013), Efficient Estimation of Word Representations in Vector Space.
14. *Miller G.* (1995), Wordnet: A lexical database for English. In Communications of the ACM.
15. *Panchenko, A., Loukachevitch, N. V., Ustalov, D., Paperno, D., Meyer, C. M., Konstantinova, N.* (2015), RUSSE: The First Workshop on Russian Semantic Similarity. In: Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue". Volume 2. RGGU, Moscow 89–105.
16. *Pennington J., Socher R., Manning, C. D.* (2014), Glove: Global vectors for word representation, Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP), Doha, Qatar, 1532–1543.
17. *Postma M., Vossen P.* (2014), What implementation and translation teach us: the case of semantic similarity measures in wordnets, Proceedings of Global WordNet Conference 2014, Tartu, Estonia, pp. 133–141.
18. *Richardson R., Smeaton A., Murphy J.* (1994), Using WordNet as a Knowledge Base for Measuring Semantic Similarity between Words, Proceedings of AICS Conference, Dublin, Ireland.
19. *Zhao S., Wang H., Liu T., Li S.* (2008), Pivot approach for extracting paraphrase patterns from bilingual corpora. In ACL: Annual Meet. of Assoc. of Computational Linguistics