

Deep Learning from Web-Scale Corpora for Better Dictionary Interfaces

Pavel Smrz

Brno University of Technology
Faculty of Information Technology
Bozotechnova 2, 612 66 Brno
Czech Republic
smrz@fit.vutbr.cz

Lubomir Otrusina

Brno University of Technology
Faculty of Information Technology
Bozotechnova 2, 612 66 Brno
Czech Republic
iotrusina@fit.vutbr.cz

Abstract

This paper explores advanced learning mechanisms – neural networks trained by the Word2Vec method – for predicting word associations. We discuss how the approach can be built into dictionary interfaces to help tip-of-the-tongue searches. We also describe our contribution to the CogALex 2014 shared task. We argue that the reverse response-stimulus word associations chosen for the shared task are only mildly related to the motivation idea of the lexical access support system. The methods employed in our contribution are briefly introduced. We present results of experiments with various parameter settings and show what improvement can be expected if more than one answer is allowed. The paper concludes with a proposal for a new collective effort to assemble real tip-of-the-tongue situation records for future, more-realistic evaluations.

1 Introduction

Human memory is fundamentally associative. To focus just on lexical access issues, it is often the case that people cannot immediately recall a word expressing a specific concept but they can give one or more words referring to concepts associated with the desired one in their minds. The failure to retrieve a word from memory, combined with partial recall and the feeling that retrieval is imminent, is generally referred to as the tip-of-the-tongue phenomenon (TOT), sometimes called *presque vu* (Brown, 1991).

Before one starts to think about automatic means supporting the lexical access, it is important to distinguish various situations in which TOT appears. First, the personal state of the language producer (writer/speaker) plays a crucial role. Fatigue or lack of attention can increase frequency of TOT situations. Specific problems come with mild cognitive impairments (incipient dementia) which is more frequent in elders. The communication mode (written or spoken language) also needs to be taken into account – it often helps to recollect an intended word if one just says associated words aloud. Consequently, people can prefer expressing the hesitation over a TOT word as a question to a family member, a friend or an automatic assistant. The spoken communication generally brings longer, more specific and detailed clues that can potentially lead to better identification of the word to be reminded. The language (mother tongue v. foreign language) and producer's familiarity and proficiency also need to be considered. Language learners would frequently associate a word with others that sound similar but are not related semantically, they could combine clues in their native language and the target one, misspell/mispronounce words, etc. Although the search across languages is not typically considered as a kind of the TOT phenomenon, we include this situation in the considered scenario.

Research prototypes of automatic assistants have to consider the above-mentioned settings and clearly identify in what types of TOT they can help. The primary decision a tool designer needs to make relates to the appropriate interface. The ultimate goal of the work described in this paper consists in integrating a TOT-aware assistants into natural user interfaces. Rather than on a desktop or tablet computer with a standard keyboard or (hand)written input, we focus on smart-phones or even wearable interfaces (smart watches, glasses), intelligent home/office infrastructure components, or robotic companions that can communicate in natural (spoken) language and that help users in their language producing tasks.

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organizers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

Although the current research deals strictly with explicitly expressed requests for a TOT-situation help, there is a possibility of automatic detection of TOT-related hesitations and immediate generation of word suggestions. In any case, the state of the art on this topic is at the beginning and these types of automatic assistants are mostly research prototypes.

To be able to evaluate the ongoing development work, the first author of this paper started to document and collect real TOT events. This includes personal experiences but also cases appearing during his communication with colleagues, family members, etc. In a relatively short time of three months, 19 documented cases were recorded. This shows that a collective effort in this area could easily lead to a new reasonably-large resource that would help to direct future research (see the concluding section). As we aim at a general TOT setting, the collected data include full descriptions of the clues, not just keyword-based TOT searches. For written-only interfaces, we provide a list of extracted keywords too. Thus, there would be the full sentence: *It is like racism but on women* (the correct answer – *discrimination*) and the set of two keywords – *racism, women* – for the written case.

Although its current limited size does not allow deriving statistically-significant results (only 3 out of 19 TOT cases can be correctly retrieved by our method if we allow 4 suggestions), the resource can be used to demonstrate crucial differences between the task of TOT- and reverse-association predictions (see the next section).

In addition to this discussion, the paper presents methods used for the stimulus-response association prediction submitted as our contribution to the CogALex 2014 shared task and their results. Section 3 introduces the methods, while Section 4 summarizes results under varying parameters. We conclude with future directions of our research and a proposal for a joint TOT-related activity.

2 Related Work

There is a long-term interest in intelligent dictionary interfaces that reflect natural lexical-access needs. Yet, advanced mechanisms of the access by meaning are rarely implemented as their integration presents significant challenges. Zock and Bilac (2004) discuss lookup mechanisms on the basis of word associations. Sinopalnikova and Smrz (2004) introduce lexical access-supporting dictionary enhancements based on various language resources – corpora, Wordnets, explanatory dictionaries and word association norms.

Free-word associations are frequently used as testing data for word relatedness experiments. Church and Hanks (1990) estimate word associations by a corpus-based association ratio. Word association thesauri or norms, representing a collection of empirical data obtained through large-scale psycholinguistic free-association tests, often define a gold standard. In particular, Zesch and Gurevych (2007) employ the University of South Florida word association, rhyme, and word fragment norms (Nelson et al., 1998) to compare characteristics of its graph representation to that of Wikipedia. Rapp (2008) experiments with associative responses to multiword stimuli on the Edinburgh Associative Thesaurus.

The CogALex 2014 shared task is very close to the experimental setting discussed in (Rapp, 2013) which also aims at computing a stimulus word leading to responses given in EAT. A fixed-window size to count word co-occurrences is used first. Log-likelihood ratios are employed to rank candidate words and products of the ranks then define the winner. Providing 7 responses (as compared to 5 in the CogALex 2014 shared task), the stimulus word is predicted with 54 % accuracy. However, only a specific subset (Kent and Rosanoff, 1910) of EAT is used which comprises 100 words. It is also not fully clear from which set of potential words target answers are chosen. This is a crucial aspect that influences accuracy. For example, Rapp (2014) took into account only primary associative responses from EAT, i. e., only 2,792 words. Obviously, it is far simpler to choose the correct answer from a limited set than from all existing words.

Other word association resources are also frequently used as test data. In addition to the Wordnet itself, they include TOEFL (Landauer and Dumais, 1997) and ESL synonym questions (Turney, 2001), RG-65 (Rubenstein and Goodenough, 1965) and WordSimilarity-353 (Finkelstein et al., 2001) test collections for degree of synonymy or SAT analogy questions (Turney et al., 2003) for the relational similarity.

Additionally, Heath et al. (2013) evaluate their association model in word guessing games (games with a purpose).

3 Free word associations v. TOT – similarities and differences

The CogALex 2014 shared task was motivated by natural lexical access but it was defined as computing reversed free-word associations. Participating automatic systems were employed to determine the most probable stimulus leading to given five most frequent responses from a free association test. For example, given words *circus*, *funny*, *nose*, *fool*, and *fun*, participating systems were supposed to compute word *clown* as the answer.

Training and test datasets came from the Edinburgh Associative Thesaurus (EAT)¹ (Kiss et al., 1972). EAT comprises about 100 associative responses given by British students for each of 8,400 stimuli. Items containing multi-word units and non-alphabetical characters were filtered out from the CogALex 2014 experimental data.

Although it has been shown that free word association norms and thesauri provide a valuable source of information for TOT-assisting (Sinopalnikova and Smrz, 2006), the two corresponding phenomena are not identical. Indeed, available data and experience clearly point out similarities but also significant differences.

Both, individual free associations as well as TOT can be full of idiosyncrasies. However, while association norms and thesauri try to present prototypical, generalized, most frequent associations, TOT assistants need to cope with personal specificity. Ideally, a system should be able to help its user remind a word given the clue *it was mentioned by Mary during our yesterday's conversation*.

Both the phenomena are also strongly culturally-dependent. Among others, this can make some resources such as large-scale corpora for particular language variants unusable. For example, let us consider the very first item from the CogALex test set – word *capable* is to be guessed as the stimulus for responses *able*, *incapable*, *brown*, *clever*, *good*. Putting aside the first two response words sharing their roots with the stimulus for a while (see the related discussion below), we come to word *brown*. This refers to *Lancelot Brown*, more commonly known as *Capability Brown* – an 18th century English landscape architect. This association is specific for the U.K. and it is hardly known to Americans. For example, the two words never collocate in the 450 million Corpus of Contemporary American English (COCA)², while *Capability Brown* is mentioned 36 times in the 100 million British National Corpus (BNC)³.

This observation led us to the question what is the overlap between two distinct word association thesauri/norms. To explore this, we compared EAT to the University of South Florida Word Association Norm (SFWAN)⁴ (Nelson et al., 1998). SFWAN consists of 5,019 normed words and their 72,176 responses. EAT and SFWAN have 3,545 stimulus terms in common. There are 11,788 words used as one or more responses in both the sets. Despite the substantial overlap of the stimulus and response sets, responses for same stimulus words in SFWAN rarely correspond to those given in EAT. Using a simple algorithm of the highest overlap among response sets, only 106 stimuli from the CogALex test set (out of 2,000) can be correctly determined from SFWAN. It can be partially explained by the cultural differences between the U.K. and the U.S.A., but also by relatively distant times of collecting/publishing the resources (1972 v. 1998), slightly different settings of the experiments and non-uniform presentation of the results. In any case, this finding casts doubts upon suitability of EAT for the shared task if no available (large) corpus data reflects the time and the setting of corresponding word association experiments (reflecting the background of students in 1972).

It can be also argued that observed associations corresponding to TOT clue words are of different nature than (reversed) free-word associations. Definitely, numbers of given clues vary, sometimes, there are two or three words only, sometimes, there are full sentences giving more than 5 keywords to associate

¹<http://www.eat.rl.ac.uk/>

²<http://corpus.byu.edu/coca/>

³<http://www.natcorp.ox.ac.uk/>

⁴<http://web.usf.edu/FreeAssociation/>

with. Spoken clues also frequently explicitly state the kind of relation of the search word to a clue (e.g., *it is an opposite to...*, *it is used for...*).

Subjects are usually instructed to give the first response in their mind to the stimulus in free word association tests. On the other hand, TOT clues are usually related to the searched word in much more subtle way. At least, it is usually enough to mention any word of the same root/stem as a candidate and the subject finds the word in TOT situations. Thus, testing free associations such as *choler-cholera*, *capable-incapable*, *misuse-abuse*, *actor-actress* is completely irrelevant for vocabulary access problems.

Native speakers have usually no problem to retrieve a word from memory if it forms a part of an idiom and the other part of the idiom is suggested. Thus, predicting either word of *tooth a nail* is not relevant for TOT situations (in any language that lexicalizes Latin *dentibus et vnguibus*). Considering lexical access in a foreign language, the reason for the same conclusion can be opposite – an idiom can be unknown to a learner so that it is not probable that a part will be given as a clue.

In languages naturally conceptualizing different parts of speech, writers or speakers always know what word category they search for. The collected data as well as intuition also suggest that TOT clues would not mix various senses of a word to be recalled. Consequently, free associations such as *stage – theatre/coach* or *March – April/Hare* have also nothing to do with TOT.

4 Methods

This section introduces methods used to compute multi-word reversed associations in our experiments. The primary method applied in the submitted system takes advantage of deep learning from web-scale corpora. To be sure that computed word associations automatically derived from large textual data cannot be matched by those resulting from a manually created resource, associations predicted by various Wordnet-based measures were also considered.

The Word2Vec technique⁵ available in Python package GenSim⁶ (Řehůřek and Sojka, 2010) was primarily utilized to predict a stimulus word from a list of most frequent responses. Word2Vec defines an efficient way to work with continuous bag-of-words (CBOW) and skip-gram (SG) architectures computing vector representations from very large data sets (Mikolov et al., 2013). The CBOW and SG approaches are both based on distributed representations of words learned by neural networks. The CBOW architecture predicts a current word based on contexts, while the SG algorithm predicts surrounding words given a current word. Mikolov et al. (2013) showed that the SG algorithm achieves better accuracies in tested cases. We have therefore applied only this architecture in our experiments. Various parameters of the training model need to be set – the dimensionality of feature vectors, the maximum distance between a current and a predicted word within a sentence or the initial learning rate. Consequently, we built various instances of the stimulus predictor varying values of the parameters. Their detailed evaluation is given in the next section.

The CogALex 2014 shared task was divided into two categories. Unrestricted systems could use any kind of data to compute results, while restricted systems were allowed only to draw on the freely available UKWaC corpus (Ferraresi et al., 2008) in order to predict word associations. We implemented systems for both the categories. Our unrestricted system employs the ClueWeb12 corpus.⁷ UKWaC comprises about 2 billion words and has size of about 30 GB (including annotations). The ClueWeb12 dataset consists of more than 733 million English web pages (collected between February and May 2012). The size of the complete ClueWeb12 data is 1.95 TB. To speed-up the process of training, only a fraction of the ClueWeb12 dataset was used to compute the Word2Vec models. It consists of about 8.7 billion words and has size of 131 GB. The ClueWeb12 data was pre-processed by removing web-page boilerplates and content duplication. The original UKWaC dataset already contains POS and lemma annotations. TreeTagger⁸ was used to produce the same input for the ClueWeb12 dataset. Some models were created from identified lemmata rather than individual tokens.

⁵<https://code.google.com/p/Word2Vec/>

⁶<http://radimrehurek.com/gensim/>

⁷<http://lemurproject.org/clueweb12/>

⁸<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

We took advantage of the nltk⁹ toolkit to experiment with Wordnet-based measures. A candidate list of all possible Wordnet-related words that could be considered as potential stimuli was computed for each of five given responses first. The word with the highest sum of similarities to all five response words was returned as the best stimulus candidate.

To populate the set of all possibly related words, standard Wordnet relations (Fellbaum, 1998) were considered – hypernyms/hyponyms, instances, holonyms/meronyms (including members and substances), attributes, entailments, causes, verb groups, see-also and similar-to relations. As the similarity measures, we used the path similarity based on the shortest path that connects the senses in the is-a (hypernym/hyponym) taxonomy, Wu-Palmer’s similarity (Wu and Palmer, 1994) based on the depth of word/sense pairs in the taxonomy and that of their Least Common Subsumer, Leacock-Chodorow’s similarity (Leacock and Chodorow, 1998) based on the shortest path that connects the senses (as above) and the maximum depth of the taxonomy in which the senses occur, Resnik’s similarity (Resnik, 1995) based on the Information Content (IC) of the least common subsumer, Jiang-Conrath’ similarity (Jiang and Conrath, 1997) based on the Information Content (IC) of the least common subsumer and that of the two input synsets and Lin’s Similarity (Lin, 1998) based on the Information Content (IC) of the least common subsumer and that of the two input synsets.

5 Evaluation

5.1 Word2Vec approach

There are various parameters to tune up when creating the models for the SG algorithm. We experimented with three of them – values *100*, *300* and *500* were tested as dimensionalities of feature vectors, values *3*, *5* and *7* for maximum distances between current and predicted words within a sentence, and the lemmatization was switched on or off. Value *word* means that original lowercased tokens were used for the computation of models, whereas value *lemma* means we used lowercased lemmata corresponding to original words. Resulting models are named accordingly: *size-window-token* (e.g., *100-3-lemma*, *500-3-word*), where *size* denotes the dimensionality of the feature vectors, *window* the maximum distance between the current and a predicted word within a sentence and *token* determines whether original words or lemmata were used for a given model. We restricted the parameters to these values mainly to cope with computational requirements. Although the Word2Vec toolkit supports multi-threaded computation, it took significant time to build all the models. For example, 60 hours in 8 threads were needed to compute the *500-7-lemma* model for the ClueWeb12 data. Although higher values for *size* and *window* parameters would probably bring better accuracies, they were not tested due to time constraints. Results for various combinations of parameters are summarized in Table 1.

EAT sometimes gives inflectional variants of words (e.g., plurals) as stimuli or responses. A strict evaluation comparing exact strings can then harm systems that do not try to match particular wordforms. To quantify the effect we compared results of our system on two versions of the test sets expanded target word lists which allow all wordforms for each target word¹⁰ and the original lists. Results are given in Table 1 in columns denoted *inflectional* in the case of the expanded lists and *non-inflectional* for the original data.

As can be seen, model *500-5-lemma* reaches the best accuracy for the unrestricted task and models *300-7-word* and *500-5-word* win in the restricted task. As only one set of results was allowed to be submitted for each task, we employed the *500-5-word* model in our submission.

Although, the CogALex 2014 shared task was defined as to predict exactly one stimulus word for five given responses, lexical access helpers can easily accommodate more suggestions. This can be evaluated by checking how frequently a gold standard stimulus appears among top *n* predicted words. Figures 1 and 2 compare results of our unrestricted and restricted systems, respectively, for up to 10 suggestions (the inflectional case). As expected, the accuracy increases with the number of candidate words taken into account. The best value of 0.4865 for the unrestricted system is reached using model *500-5-lemma*, while the best accuracy of 0.4575 for the restricted system comes from model *500-7-word*.

⁹<http://www.nltk.org/>

¹⁰Provided by Michael Flor and Beata Beigman Klebanov (ETS Princeton).

model	non-inflectional		inflectional	
	unrestricted	restricted	unrestricted	restricted
100-3-lemma	0.11	0.083	0.1215	0.0865
100-3-word	0.1005	0.098	0.11	0.1015
100-5-lemma	0.1055	0.1	0.1165	0.1045
100-5-word	0.1115	0.1165	0.1195	0.1225
100-7-lemma	0.1235	0.1035	0.1345	0.108
100-7-word	0.112	0.1265	0.1235	0.137
300-3-lemma	0.178	0.1395	0.1945	0.1475
300-3-word	0.1605	0.157	0.1705	0.163
300-5-lemma	0.179	0.1525	0.196	0.161
300-5-word	0.175	0.183	0.19	0.193
300-7-lemma	0.1875	0.158	0.206	0.167
300-7-word	0.17	0.195	0.1885	0.207
500-3-lemma	0.188	0.1395	0.203	0.1465
500-3-word	0.174	0.176	0.1845	0.1845
500-5-lemma	0.1975	0.161	0.219	0.1685
500-5-word	0.1795	0.195	0.1955	0.2065
500-7-lemma	0.193	0.169	0.2075	0.1795
500-7-word	0.191	0.194	0.209	0.2085

Table 1: Accuracies of Word2Vec-based methods with varying parameters

Together with their original Word2Vec implementation, Mikolov et al. (2013) made available also word vectors resulting from training on a part of the Google News dataset (consisting of 100 billion words). The model contains 300-dimensional vectors for 3 millions of words and phrases (no lemmatization was performed). We repeated CogALex 2014 shared task experiments with this pre-trained model as well. The resulting accuracy was 0.1375. The lower value is probably caused by the fact that the model is trained on the specific dataset with different pre-processing.

5.2 Wordnet-based measures

The Wordnet-based approach was evaluated in the same way as the Word2Vec one. Result for all six similarity measures are listed in Table 2. As in the previous case, accuracies for top n ($1 \leq n \leq 10$) predicted responses are considered. The best performing Wordnet similarity measure for the task showed to be Lin’s similarity based on the Information Content of the least common subsumer and that of the two input synsets. Yet, the best values are far from accuracies of the Word2Vec-based methods, especially when only few predicted responses are allowed. This confirms our hypothesis that approaches deriving their lexical knowledge from large textual corpora overcome those based only on Wordnet.

6 Conclusions and future directions

The CogALex 2014 shared task focused on computing reversed multi-word response-stimulus relations extracted from the Edinburgh Association Thesaurus. We showed that this setting is only weakly related to computer-aided lexical access problems, namely to the tip-of-the-tongue phenomenon.

The submitted results were obtained with a system based on the Word2Vec distributional similarity model. Best of the implemented systems reaches accuracy of 0.1975 when trained on a subset of the ClueWeb12 dataset. Unfortunately, in time of writing this paper, official results of other teams are not published. Hence, no comparison with other participants could be included.

Section 2 also mentions our experience in collecting real TOT data. We believe that a collective effort could lead to a much larger resource better reflecting nature of the TOT phenomenon. We propose to establish a task force aiming at this goal. During discussions at the workshop, we could focus on actual

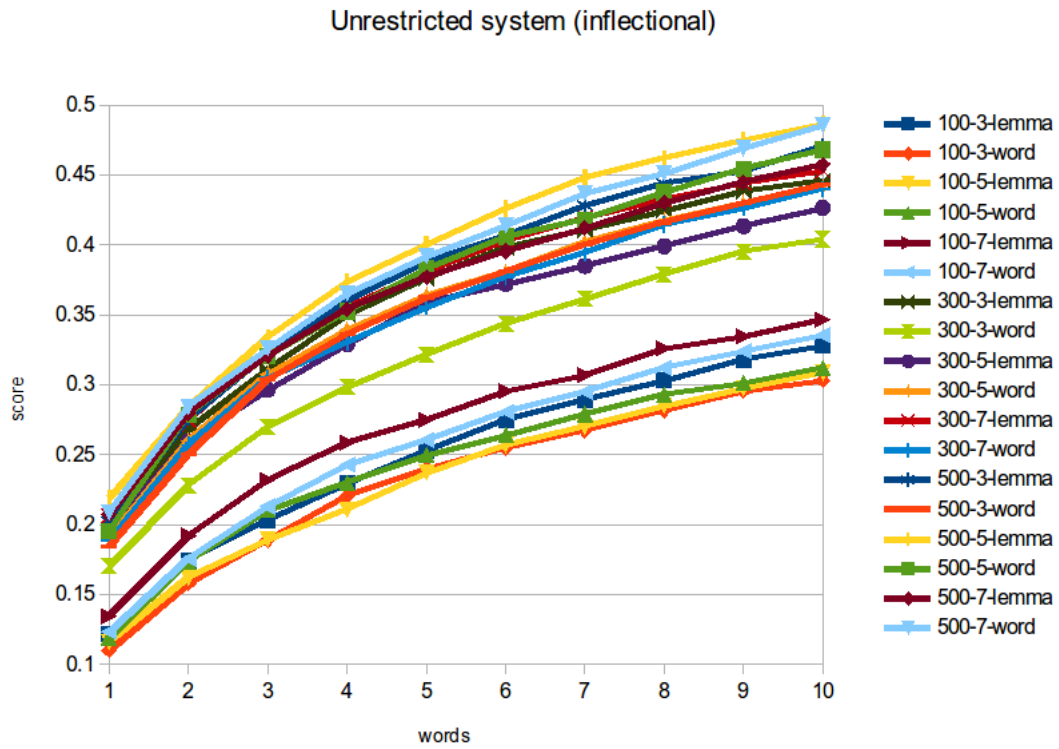


Figure 1: Accuracies growing with the number of suggestions for the unrestricted system

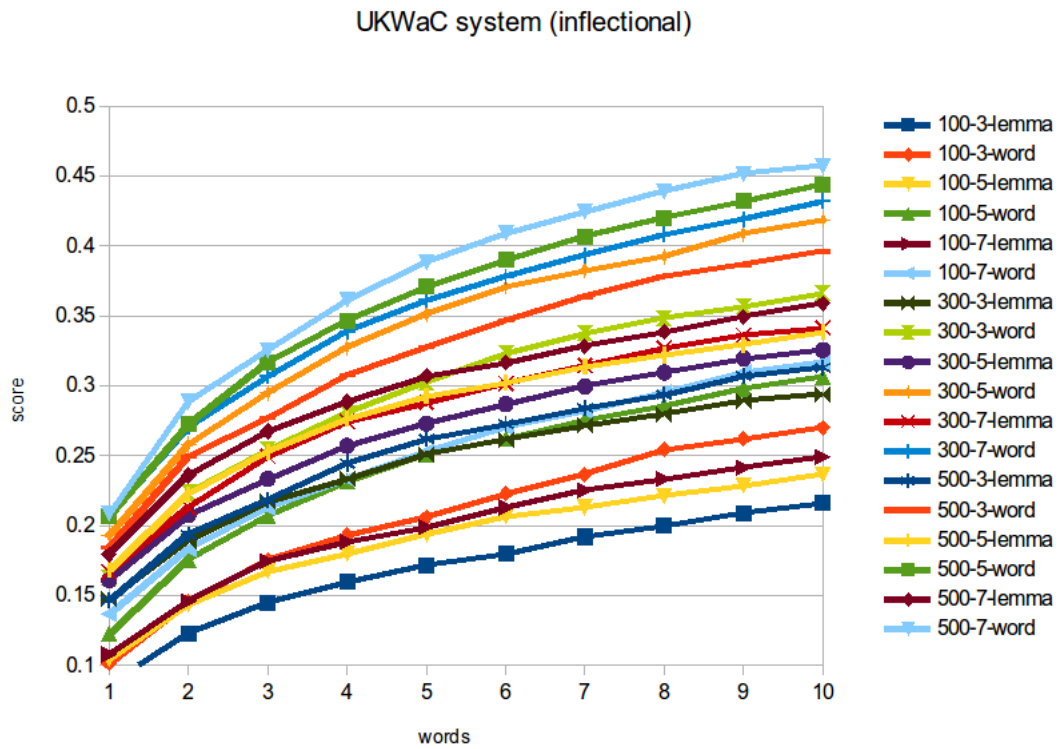


Figure 2: Accuracies growing with the number of suggestions for the restricted system

sim.	top n predicted responses are considered									
	1	2	3	4	5	6	7	8	9	10
path	0.0085	0.0175	0.024	0.0325	0.0395	0.044	0.0525	0.057	0.063	0.066
wup	0.012	0.03	0.0455	0.057	0.068	0.0745	0.0875	0.095	0.1025	0.1075
lch	0.003	0.0085	0.015	0.0195	0.0215	0.028	0.034	0.0375	0.04	0.042
res	0.008	0.0205	0.031	0.0435	0.048	0.061	0.0715	0.0785	0.0845	0.09
jcn	0.0135	0.029	0.042	0.0575	0.065	0.0825	0.098	0.1105	0.12	0.1295
lin	0.022	0.044	0.068	0.091	0.1045	0.1265	0.1485	0.1675	0.1805	0.1955

Table 2: Results of Wordnet-based methods (*path* stands for the path similarity, *wup* for Wu-Palmer’s similarity, *lch* for Leacock-Chodorow’s similarity, *res* for the Resnik’s similarity, *jcn* for the Jiang-Conrath’s similarity and *lin* for Lin’s similarity).

procedures and technical support means (through a web-based system) to build the resource within the next year. The collected dataset could then be used for future shared tasks in the domain.

Acknowledgements

The research leading to these results has received support from the IT4Innovations Centre of Excellence project, Reg. No. CZ.1.05/1.1.00/02.0070, supported by Operational Programme “Research and Development for Innovations” funded by Structural Funds of the European Union and the state budget of the Czech Republic.

References

- Alan S. Brown. 1991. A review of the tip-of-the-tongue experience. *Psychological Bulletin*, 109(2):204–223, March.
- Kenneth Ward Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29, March.
- Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA; London, May. ISBN 978-0-262-06197-1.
- Adriano Ferraresi, Eros Zanchetta, Marco Baroni, and Silvia Bernardini. 2008. Introducing and evaluating ukwac, a very large web-derived corpus of english. In *Proceedings of the 4th Web as Corpus Workshop (WAC-4) Can we beat Google*, pages 47–54.
- Lev Finkelstein, Gabrilovich Evgenly, Matias Yossi, Rivlin Ehud, Solan Zach, Wolfman Gadi, and Ruppim Eytan. 2001. Placing search in context: the concept revisited. In *Proceedings of the Tenth International World Wide Web Conference*.
- Derrall Heath, David Norton, Eric K. Ringger, and Dan Ventura. 2013. Semantic models as a combination of free association norms and corpus-based correlations. In *ICSC*, pages 48–55. IEEE.
- J.J. Jiang and D.W. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. *Arxiv preprint cmp-lg/9709008*.
- Grace Helen Kent and Aaron Joshua Rosanoff. 1910. *A study of association in insanity*. American Journal of Insanity.
- GR Kiss, Christine A Armstrong, and R Milroy. 1972. *An associative thesaurus of English*. Medical Research Council, Speech and Communication Unit, University of Edinburgh, Scotland.
- T. K. Landauer and S. T. Dumais. 1997. A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211–240.
- C. Leacock and M. Chodorow. 1998. Combining local context and WordNet similarity for word sense identification. *WordNet: An electronic lexical database*, 49(2):265–283.

- D. Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*, volume 1, pages 296–304. Citeseer.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.
- D. L. Nelson, McEvoy, C. L., and T. A. Schreiber. 1998. The University of South Florida word association, rhyme, and word fragment norms. <http://w3.usf.edu/FreeAssociation/>.
- Reinhard Rapp. 2008. The computation of associative responses to multiword stimuli. In *Proceedings of the Workshop on Cognitive Aspects of the Lexicon*, COGALEX '08, pages 102–109. Association for Computational Linguistics.
- Reinhard Rapp. 2013. From stimulus to associations and back. *Natural Language Processing and Cognitive Science*, page 78.
- Reinhard Rapp. 2014. Using word association norms to measure corpus representativeness. In *Computational Linguistics and Intelligent Text Processing*, pages 1–13. Springer.
- Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May. ELRA. <http://is.muni.cz/publication/884893/en>.
- P. Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. *Arxiv preprint cmp-lg/9511007*.
- Herbert Rubenstein and John B. Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM*, 8(10):627–633.
- Anna Sinopalnikova and Pavel Smrz. 2004. Word association norms as a unique supplement of traditional language resources. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, pages 1557–1561, Lisbon. European Language Resources Association.
- Anna Sinopalnikova and Pavel Smrz. 2006. Knowing a word vs. accessing a word: Wordnet and word association norms as interfaces to electronic dictionaries. In *Proceedings of the Third International WordNet Conference*, pages 265–272.
- Peter D. Turney, Michael L. Littman, Jeffrey Bigham, and Victor Shnayder. 2003. Combining independent modules to solve multiple-choice synonym and analogy problems. *CoRR*, cs.CL/0309035.
- Peter D. Turney. 2001. Mining the web for synonyms: Pmi-ir versus lsa on toefl. In *Proceedings of the 12th European Conference on Machine Learning*, EMCL '01, pages 491–502. Springer-Verlag.
- Z. Wu and M. Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138. Association for Computational Linguistics.
- Torsten Zesch and Iryna Gurevych. 2007. Analysis of the wikipedia category graph for nlp applications. In *Proceedings of the TextGraphs-2 Workshop (NAACL-HLT)*.
- Michael Zock and Slaven Bilac. 2004. Word lookup on the basis of associations: From an idea to a roadmap. In *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries*, ElectricDict '04, pages 29–35. Association for Computational Linguistics.