Using BERT to Explore Lexical Semantic Change of Prepositions

Liudmila Radchankava[1] and Vasily Konovalov[2,3]
[1]Johannes Gutenberg University Mainz
[2]AIRI, [3]MIRAI
luradcha@uni-mainz.de

Abstract

This paper presents a semi-supervised approach to explore the lexical semantic change of English prepositions using the BERT language model. We adopt a method that leverages BERT to obtain representations of prepositions, train a model on these representations, and use this model to measure the lexical semantic changes of the prepositions. We show that BERT embeddings can be used for preposition-sense disambiguation. Then, by applying the trained model to sentences extracted from the Corpus of Historical American English (COHA), we detect a variety of lexical semantic changes over time for English prepositions.

1 Introduction

Many linguists distinguish between lexical categories and functional categories of words. While functional categories have mostly grammatical meanings, lexical categories have a more obvious descriptive content. In the case of prepositions, linguists argue that there are two subclasses: function prepositions and content prepositions. Accordingly, some assume that the class of prepositions includes both grammatical and lexical elements, analogous to the distinction between content verbs and function verbs (auxiliaries) (Foucambert and Zuniga, 2012).

It has also been suggested that prepositions are a hybrid between the two categories depending on their usage. Therefore, the prepositions cluster can be located between grammatical words (like determiners) on the one hand, and lexical words (nouns, verbs, adjectives and adverbs) on the other. For example, the English preposition *of* is grammatical, whereas *off* is lexical (Boye and Harder, 2012).

It is broadly known that prepositions are a highly polysemic class (association of one word with two or more distinct meanings). The preposition *on*, for example, is originally a spatial adposition (*the book is on the table*), it may express temporal relations (*it was on 10 March*), causal relations (*she read the book on my suggestion*), or relations of manner (*she spoke on a rising tone*) (Stefanowitsch et al., 2020).

Using the correct sense of a preposition is important for understanding the meaning of the utterance (Baldwin et al., 2009). This task becomes even more challenging because of the scarcity of the training dataset. Prepositions have not received adequate attention in the NLP community. Even in the Penn Treebank, only two types of prepositions are recognized (*in* (locative, temporal, and manner) and *to* (direction)) (O'Hara, 2005).

Bidirectional Encoder Representations from Transformers (BERT) are a language representation model. Unlike other models, BERT is designed to pre-train deep bidirectional representations from an unlabeled text by jointly conditioning on both left and right contexts in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, including text classification, sequence classification, and question answering, without significant task-specific architecture modifications. Google AI released multiple versions of BERT, such as multilingual BERT, English BERT, and Chinese BERT (Google-Research, 2018; Devlin et al., 2019).

In this work we apply BERT to the task of preposition-sense disambiguation and measure lexico-semantic change of prepositions. Our contributions are:

1. We show that despite the mixed nature of prepositions, BERT-based contextualized word embeddings used by the kNN classifier can address preposition-sense disambiguation and achieve reasonable performance in comparison to the previous systems that leverage external resources. Also, we show that mul-

The Proceedings for the 6th International Workshop on Computational Approaches to Language Change (LChange'26), pages 124–130

tilingual BERT performance falls behind the language-specific English base BERT by a small margin.

2. We propose a distance-to-labeled-exemplars signal derived from BERT embeddings as a proxy for diachronic contextual divergence in prepositions.

3. We provide empirical evidence that, under a distance-based proxy derived from contextual embeddings, compound prepositions show clearer diachronic shifts in usage contexts than simple prepositions, whose developments are typically more gradual and harder to detect automatically.

Even though our study was applied to English prepositions, the same methods can be applied to any language using multilingual BERT provided an appropriate preposition dataset.

2 Related Work

Exploring usage types of prepositions is a part of the Word Sense Disambiguation (WSD) task. The goal of WSD is to identify the correct sense of the word usage from a fixed set of sense identifiers. To train and evaluate WSD systems, many shared task datasets have been published in the SemEval workshop series.

Previous approaches to preposition-sense disambiguation (PSD) were based on external knowledge resources as WordNet (extracting synonyms and hypernyms) including extensive feature engineering. Instead of using BOW that focuses solely on the words within a fixed window size, they focus on elements that are related via the phrase structure (Tratz and Hovy, 2009).

Also, multilingual data can be used to improve the accuracy of PSD. The approach is to pretrain an LSTM encoder to predict the translation of a preposition, and then to incorporate the pre-trained encoder as a component in a supervised classification system, as well as fine-tune it for the task. The multilingual signals consistently improve the results on preposition-sense dataset (Gonen and Goldberg, 2016).

It has been shown that BERT embeddings can be utilized directly to approach the WSD task due to their property of providing distinct vector representations for the same token depending on its context. To learn the semantic capabilities of CWEs, a simple yet interpretable kNN approach has been employed. However, the approach has only been tested on nouns (Wiedemann et al., 2019).

Moreover, BERT has been found useful in exploring over-time development of a preposition (only of *about*). The study focused on its three most frequent senses. It demonstrated that unsupervised BERT-embeddings can identify the usage types of *about* in English. It formulated three expected functional-semantic developments of *about* based on linguistic literature, and assessed whether BERT-embeddings can be used in combination with Entropy Difference measures to (semi-) automatically detect these changes in sentences between 1810 and 2009 taken from the Corpus of Historical American English (COHA) (Fonteyn, 2020).

The multilingual BERT was trained on Wikipedia for 104 languages and uses a shared cross-language BPE vocabulary. Its cross-lingual transferability was demonstrated on sequential tasks like NER and POS (Pires et al., 2019), and in multi-tasking mode (Karpov and Konovalov, 2023).

3 Datasets

To train the model we used The Preposition Project (TPP) dataset, which was the dataset for SemEval 2007 Task 6: Word-Sense Disambiguation of Prepositions. This corpus covers 34 prepositions with 16,557 training and 8,096 test sentences, each containing a single preposition example. The sentences were extracted from the FrameNet database, based mostly on the British National Corpus (designed to represent a wide cross-section of British English, both spoken and written, from the late twentieth century). Each preposition has a different set of possible senses. The prepositions were labeled in accordance with the Oxford Dictionary of English (ODE). We used the original split to train and test sets (Litkowski and Hargraves, 2007).

4 Experimental Setup

For PSD, we rely on contextualized representations of prepositions. We leverage the BERT language model to compute preposition representations for each occurrence in an utterance. Then we train the model given the contextualized preposition representation and corresponding semantic classes. Having trained the model we label all the given preposition use cases along the temporal axis. Then we detect and analyze the occurrences associated with larger distances to the nearest neighbor.

4.1 BERT Language Model

We produce usage representations using the BERT language model, a multi-layer bidirectional Transformer encoder trained on masked token prediction and next sentence prediction. We use the original English BERT base model `bert-base-uncased` with 12 layers, 768 hidden dimensions, and 110M parameters. To measure the effect of the number of parameters we test the large BERT model (`bert-large-uncased`). Also, we use the multilingual BERT model (`bert-base-multilingual-uncased`) to measure to what extent multilingual BERT can be applied to language-specific PSD.

Given a word of interest w and a context of occurrence $c = (v_1, ..., v_n)$, with $w = v_i$, we extract the preposition embeddings from the last four hidden layers of the encoder and average them (Savkin et al., 2024). It appears that prepositions are frequent enough tokens and they are never divided into sub-tokens in our settings.

4.2 Nearest-Neighbor Classifier

We apply a non-parametric nearest-neighbor classifier (kNN) to label the preposition senses. The kNN classification algorithm (Cover and Hart, 1967) assigns a plurality vote of a sample's nearest labeled neighbors. We select kNN classifier because its effectiveness for PSD is mainly based on the representation abilities of BERT. Moreover, it can directly lead to the training example that sets a certain classifier decision. In the most simple case, one-nearest neighbor, it predicts the label from the nearest training instance by some defined distance metric. As distance measure for kNN, we use the Euclidean distance of the CWE vectors.

Our approach considers only senses for a target word that has been observed during training. This approach has been named as localized nearest neighbor word sense disambiguation (Wiedemann et al., 2019).

4.3 Detecting Semantic Change

The TPP prepositions can be divided between simple and compound prepositions (from diachronic point of view). The simple category contains the most ancient prepositions, e.g. *in*, *on*, *of*, *by*, *with* etc. The compound prepositions, however, contains the majority of prepositions (cf. *about, across, along, around, aside, beside, onto, into, aback, above, after, again, apart, astray, asunder, athwart, before, behind, below, between* etc.) (Lindstromberg, 2010). Simple prepositions could be viewed as grammatical items (they fulfill the valency of a verb e.g. to belong to smb, to depend on smth/smb, to fear for smb), while compound prepositions are still between two poles and can have both grammatical and lexical characteristics (Lehmann, 2015). It can be assumed that simple prepositions tend to exhibit more gradual, highly polysemous, and context-dependent semantic developments, which are harder to capture using distributional methods, whereas compound prepositions are more likely to undergo discretely identifiable lexico-semantic changes (Lehmann, 2015; Rhee, 2002). This does not imply that simple prepositions do not undergo semantic change. Rather, their changes are often incremental and organized in radial semantic structures, making them less salient for automatic detection based on contextual embedding distances. We analyze five representative prepositions from each group (*in*, *on*, *at*, *with*, *of*) and five representative compound prepositions (*into*, *onto*, *beside*, *above*, *about*). Accordingly, cases where OED-documented senses predate the benchmark period do not contradict our findings, as the proposed proxy captures shifts in usage contexts rather than the emergence of entirely new dictionary senses.

For each preposition, we sample 2,000 sentences from the Corpus of Historical American English (COHA). The corpus is balanced by genre across decades; sentences were sampled uniformly from COHA without additional temporal stratification. COHA is the largest structured corpus of historical English. It contains more than 475 million words of text from the 1820s-2010s (which makes it 50-100 times as large as other comparable historical corpora of English) and the corpus is balanced by genre decade by decade (Davies, 2012).

We apply BERT to extract CWE of prepositions from the sampled sentences, then we measure the distance between the preposition representation from the sample and its top-1 neighbor from TPP. Larger distances correspond to lower classifier confidence, which we interpret as a proxy for contextual divergence and potential diachronic change.

As the last step, we compute Spearman's rank correlation between the usage year and the distance to the top-1 neighbor. Spearman's ρ tests for a monotonic association without assuming normality. We report two-sided p-values and treat correlations with $p \leq 0.05$ as statistically significant.

5 Experimental Results

We conduct two types of experiments. We calculate the performance of the kNN model on the SemEval dataset to justify that our settings can be applied to solve PSD despite mixed lexico-grammatical origins of the prepositions. In our second experiment, we use the trained model to label sampled prepositions from COHA over time. Furthermore, we identify the old-fashioned usages by leveraging the classifier's confidence (the distance to the nearest neighbor).

5.1 Nearest-Neighbor Model

In the SemEval corpus, each preposition has a different set of senses, and the natural approach is to learn a different model for each preposition.

To obtain a more robust nearest-neighbor classification, we optimize for the k parameter. The optimization is performed on the development set (80%/20% split of the original training set).

Table 1 shows results of k optimization for the English base BERT model, large BERT model, multilingual BERT. The best result, with an accuracy of 80.2, is achieved for the base BERT model and $k = 5$. Surprisingly, the base BERT model outperforms the large BERT model by a small margin, whereas the large BERT model has a comparable performance with the multilingual BERT model for $k > 6$.

Table 2 compares results of the base BERT model on the test set with those of previous systems. The First Sense baseline selects the first sense of each preposition as the answer (under the assumption that the senses are organized somewhat according to prominence). The Freq Sense baseline selects the most frequent sense from the training set.

The KU system uses a statistical language model based on a large unannotated corpus. The model is used to evaluate the likelihood of various substitutes for a word in a given context. These likelihoods are then used to determine the best sense for the word in novel contexts (Yuret, 2007).

The MELB-YB system achieved the highest result out of the three participating systems in the SemEval 2007 shared task. The approach leveraged features such as POS tags and WordNet-based features, and also high-level features (semantic role tags), using a word window of up to seven words, in a Maximum Entropy classifier (Ye and Baldwin, 2007).

Model	*base-BERT*	*large-BERT*	*M-BERT*
k=1	79.4	76.9	73.9
k=2	77.3	75.3	72.3
k=3	79.7	77.5	75.4
k=4	79.6	77.5	75.5
k=5	**80.2**	77.4	75.5
k=6	80	77.2	75.4
k=7	80	77.1	76
k=8	79.6	76.9	75.4
k=9	79.5	76.8	75.3
k=10	79.5	76.7	75.2

Table 1: Optimizing for k parameter.

Model	Accuracy
Our system (`bert-base-uncased`)	79.8
LSTM+ (Gonen and Goldberg, 2016)	**81.7**
MELB-YB (Ye and Baldwin, 2007)	69.3
KU (Yuret, 2007)	54.7
IRST-BP (Popescu et al., 2007)	49.6
Most Frequent Sense	39.6
First Sense	28.9

Table 2: The accuracy on the test set of the TPP dataset (the SemEval corpus).

The approach that pretrains an LSTM encoder to predict the translation of a preposition, incorporating the pre-trained encoder as a component in a supervised classification system and fine-tuning it for the task, achieves 81.7 (Gonen and Goldberg, 2016).

Our kNN base BERT model achieves an accuracy of 79.8. It outperforms all SemEval 2007 participants, though it falls short of the LSTM-based model that leverages external multilingual data. In contrast, our models do not include any external knowledge, based solely on CWE of BERT.

The goal of this comparison is not to achieve the state-of-the-art performance but to show that BERT CWE along with the kNN classifier can be used to solve PSD, which is different from noun or verb sense disambiguation due to the mixed lexico-grammatical roles of prepositions in sentences.

5.2 Detecting Semantic Change

In our second experiment, we label all the utterance samples from COHA with the corresponding preposition classifier and assign a distance to the nearest neighbor (a measure of confidence). Then we calculate the correlation between these distances and the usage year taken from COHA. Table 3 shows the Spearman correlation coefficient with the corresponding p-value. According to the correlation results we have two groups. The group of simple prepositions (*on*, *at*, *in*, *of*, *with*) either shows low correlation close to 0, or has a high p-value, which means that null hypothesis is correct (no correlation). This pattern is consistent with the idea that, under our distance-based proxy, simple prepositions show weaker or less consistently detectable diachronic trends, while their changes may be more gradual and harder to capture automatically (Lehmann, 2015; Rhee, 2002). However, all compound prepositions except for *into* have a negative correlation, suggesting that earlier attestations tend to be associated with larger distances (lower confidence) under our proxy.

Preposition	Correlation
on	$0.06(3.8 \cdot 10^{-3})$
in	$-0.005(\mathrm{NS})$
at	$-0.06(2.4 \cdot 10^{-3})$
of	$0.05(\mathrm{NS})$
with	$0.05(4.1 \cdot 10^{-2})$
into	$0.075(4.91 \cdot 10^{-12})$
above	$-0.13(5.24 \cdot 10^{-37})$
onto	$-0.10(3.19 \cdot 10^{-11})$
beside	$-0.22(4.56 \cdot 10^{-43})$
about	$-0.13(1.18 \cdot 10^{-35})$

Table 3: Spearman correlation coefficient with corresponding p-value; NS indicates $p > 0.05$.

onto The preposition *onto* is used with verbs to express movement on or to a particular place or position: *Move the books onto the second shelf*. The preposition *onto* is also used to show that something faces in a particular direction: *The window looked out onto the terrace*. Larger distances to the nearest neighbor for *onto* were mainly observed in sociolectal (non-standard register) contexts.

I wuz a settin onto a rockincheer, and Hanner Ann wuz on my knee. (1865)
... the stoppin to pick up them ez coodent stick onto ther flyin steeds, I hed no difficulty in outrunnin em. (1866)
... vagary uv the mind, wich, wen loosed from its clay, sores off onto its own hook, without any restraint. Is the giant Republican actually dead. (1869)

above According to OED the preposition *above* expresses position in or movement to a place that is higher: *The water came above our knees*; more than something; greater in number, level or age than somebody/something: *Inflation is above 6%*. Larger distances to the nearest neighbor were caused by a slightly outdated preposition usage: *the splintered rapiers of their opponents appeared a speedy omen of the enforcement of the threats of utter annihilation, that was hoarsely shrieked above the tumult of the contest.* Here *above* is used as in examples "shout above the noise" or rather "shout over the noise". The next interesting usage, "not being above four feet" for "being less than four feet": *south shore was much more elevated than that near the river, which had here extremely low banks, the water in the stream not being above four feet below them*

beside Preposition *beside* is used in a sense of next to or at the side of somebody/something: *He sat beside her all night*; compared with somebody/something: *My painting looks childish beside yours*. Larger distances to the nearest neighbor were often observed in contexts with a nearby pronoun, which may reflect contextual similarity patterns in the TPP benchmark rather than the emergence of a new sense.

scene was going on in the cell of the destined victim. His daughter kneeled beside him at daylight in his prison. She had cheered his solitude with the sunshine. (1832)
sister Bel has reason to be thankful, " said Catharine, who was close beside her sister, " for your teaching her name so familiarly to the river-gods. (1831)

into According to OED the preposition *into* expresses motion from without to a point within limits of space, time, condition, circumstance, etc.; the motion which results in the position expressed by in, or which is directed towards that position: *Come into the house.*; in the direction of something: *Speak clearly into the microphone.* Preposition *into* occurs most frequently among compound prepositions and seems to be better integrated into the grammatical system. Accordingly, *into* shows a comparatively weak trend under our proxy (despite statistical significance), suggesting a relatively stable usage pattern. However, there are some outdated examples:

[...]in nations; and he was convinced that, before vice could be thus exalted into custom , there must exist in the community which would tolerate such an institution (1816)
It is possible that these circumstances may have occasionally betrayed me into intemperances of expression which I did not intend: it is certain, that I (1817)

5.2.1 Limitations and Interpretation

The distance to the nearest labeled neighbor should not be interpreted as a direct measure of semantic innovation. Instead, it serves as a proxy for contextual divergence from a fixed, contemporary sense inventory derived from the SemEval-2007 dataset. Importantly, the SemEval-2007 sense inventory is substantially imbalanced: frequent senses are more densely represented in embedding space and therefore tend to yield smaller nearest-neighbor distances. For this reason, we interpret distances comparatively over time within the same preposition, rather than as absolute indicators of novelty. Moreover, increasing distances may reflect not only semantic reanalysis but also contextual expansion, stylistic drift, register change, or corpus composition effects. Consequently, the proposed method is intended as an exploratory tool for identifying diachronic shifts in usage patterns rather than a definitive test of lexical semantic change in the lexicographic sense.

6 Conclusion

Some of the detected effects, particularly for prepositions such as beside, may reflect contextual expansion rather than the emergence of entirely new senses as documented in historical dictionaries such as the OED. Our distance-based measure is sensitive to shifts in usage contexts, which can manifest as increasing semantic conventionalization or stylistic change. In this sense, our results capture a broader notion of language change, encompassing both semantic reanalysis and contextual expansion. In this work, we measured to what extent BERT-based contextualized word embeddings can address preposition-sense disambiguation. To test their capabilities to distinguish different preposition senses we used a kNN classifier. For our experiments, we used TPP datasets from SemEval 2007 task 6. We can conclude that the kNN model trained on BERT embeddings achieves reasonable performance, despite the mixed lexico-grammatical origins of prepositions. Also, by using the trained model we observe a lexical semantic change in compound prepositions and analyze it ad hoc.

References

Timothy Baldwin, Valia Kordoni, and Aline Villavicencio. 2009. Prepositions in applications: A survey and introduction to the special issue. *Comput. Linguistics*, 35(2):119–149.

Kasper Boye and Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language*, 88:1–44.

Thomas M. Cover and Peter E. Hart. 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory*, 13(1):21–27.

Mark Davies. 2012. The 400 million word corpus of historical american english (1810–2009). pages 231–262.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Lauren Fonteyn. 2020. What about grammar? using BERT embeddings to explore functional-semantic shifts of semi-lexical and grammatical constructions. In *Proceedings of the Workshop on Computational Humanities Research (CHR 2020), Amsterdam, The Netherlands, November 18-20, 2020*, volume 2723 of *CEUR Workshop Proceedings*, pages 257–268. CEUR-WS.org.

Denis Foucambert and Michael Zuniga. 2012. Effects of grammatical categories on letter detection in continuous text. *Journal of psycholinguistic research*, 41:33–49.

Hila Gonen and Yoav Goldberg. 2016. Semi supervised preposition-sense disambiguation using multilingual data. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 2718–2729. ACL.

Google-Research. 2018. Tensorflow code and pretrained models for bert. https://github.com/google-research/bert.

Dmitry Karpov and Vasily Konovalov. 2023. Knowledge transfer between tasks and languages in the multi-task encoder-agnostic transformer-based models. In *Computational Linguistics and Intellectual Technologies*, volume 2023.

Christian Lehmann. 2015. *Thoughts on grammaticalization*. Number 1 in Classics in Linguistics. Language Science Press, Berlin.

Seth Lindstromberg. 2010. English prepositions explained: Revised edition. pages 1–273.

Kenneth C. Litkowski and Orin Hargraves. 2007. Semeval-2007 task 06: Word-sense disambiguation of prepositions. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval@ACL 2007, Prague, Czech Republic, June 23-24, 2007*, pages 24–29. The Association for Computer Linguistics.

Thomas P O'Hara. 2005. *Empirical acquisition of conceptual distinctions via dictionary definitions*. Ph.D. thesis, Citeseer.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert? In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4996–5001. Association for Computational Linguistics.

Octavian Popescu, Sara Tonelli, and Emanuele Pianta. 2007. IRST-BP: preposition disambiguation based on chain clarifying relationships contexts. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval@ACL 2007, Prague, Czech Republic, June 23-24, 2007*, pages 191–194. The Association for Computer Linguistics.

Seongha Rhee. 2002. Semantic changes of english preposition against: A grammaticalization perspective.

Maksim Savkin, Anastasia Voznyuk, Fedor Ignatov, Anna Korzanova, Dmitry Karpov, Alexander Popov, and Vasily Konovalov. 2024. DeepPavlov 1.0: Your gateway to advanced NLP models backed by transformers and transfer learning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 465–474, Miami, Florida, USA. Association for Computational Linguistics.

Anatol Stefanowitsch, Elena Smirnova, and Matthias Hüning. 2020. *2 Complex adpositions in three West Germanic Languages: German, Dutch, and English*, pages 65–138. De Gruyter Mouton, Berlin, Boston.

Stephen Tratz and Dirk Hovy. 2009. Disambiguation of preposition sense using linguistically motivated features. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, May 31 - June 5, 2009, Boulder, Colorado, USA, Student Research Workshop and Doctoral Consortium*, pages 96–100. The Association for Computational Linguistics.

Gregor Wiedemann, Steffen Remus, Avi Chawla, and Chris Biemann. 2019. Does BERT make any sense? interpretable word sense disambiguation with contextualized embeddings. In *Proceedings of the 15th Conference on Natural Language Processing, KONVENS 2019, Erlangen, Germany, October 9-11, 2019*.

Patrick Ye and Timothy Baldwin. 2007. MELB-YB: preposition sense disambiguation using rich semantic features. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval@ACL 2007, Prague, Czech Republic, June 23-24, 2007*, pages 241–244. The Association for Computer Linguistics.

Deniz Yuret. 2007. KU: word sense disambiguation by substitution. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval@ACL 2007, Prague, Czech Republic, June 23-24, 2007*, pages 207–214. The Association for Computer Linguistics.