UWB @ DIACR-Ita: Lexical Semantic Change Detection with CCA and Orthogonal Transformation

Ondřej Pražák, Pavel Přibáň and Stephen Taylor

OpenEdition books

http://books.openedition.org

UWB @ DIACR-Ita: Lexical Semantic Change Detection with CCA and Orthogonal Transformation

Ondřej Pražák[* 1,2], **Pavel Přibáň**[* 1,2], and **Stephen Taylor**[* 2]

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

Abstract

In this paper, we describe our method for detection of lexical semantic change (i.e., word sense changes over time) for the DIACR-Ita shared task, where we ranked 1st. We examine semantic differences between specific words in two Italian corpora, chosen from different time periods. Our method is fully unsupervised and language independent. It consists of preparing a semantic vector space for each corpus, earlier and later. Then we compute a linear transformation between earlier and later spaces, using CCA and Orthogonal Transformation. Finally, we measure the cosines between the transformed vectors.

1 Introduction

Language evolves with time. New words appear, old words fall out of use, and the meanings of some words shift. There are changes in topics, syntax, and presentation structure. Reading the natural philosophy musings of aristocratic amateurs from the eighteenth century, and comparing with a monograph from the nineteenth century, or a medical study from the twentieth century, we can observe differences in many dimensions, some of which need a deep historical background to study. Changes in word senses are both a visible and a tractable part of language evolution.

Computational methods for researching the stories of words have the potential of helping us understand this small corner of linguistic evolution. The tools for measuring these diachronic semantic shifts might also be useful for measuring whether the same word is used in different ways in synchronic documents. The task of finding word sense changes over time is called di-achronic *Lexical Semantic Change (LSC)* detection. The task is getting more attention in recent years (Hamilton et al., 2016b; Schlechtweg et al., 2017; Schlechtweg et al., 2020). There is also the *synchronic LSC* task, which aims to identify domain-specific changes of word senses compared to general-language usage (Schlechtweg et al., 2019).

1.1 Related Work

Tahmasebi et al. (2018) provide a comprehensive survey of techniques for the *LSC* task, as do Kutuzov et al. (2018). Schlechtweg et al. (2019) evaluate available approaches for *LSC* detection using the *DURel* dataset (Schlechtweg et al., 2018). Schlechtweg et al. (2020) present results of the first shared task that addresses the LSC problem and provide an evaluation dataset that was manually annotated for four languages.

According to Schlechtweg et al. (2019), there are three main types of approaches. (1) Semantic vector spaces approaches (Gulordava and Baroni, 2011; Eger and Mehler, 2016; Hamilton et al., 2016a; Hamilton et al., 2016b; Rosenfeld and Erk, 2018; Pražák et al., 2020) represent each word with two vectors for two different time periods. The change of meaning is then measured by some distance (usually by the cosine distance) between the two vectors. (2) Topic modeling approaches (Bamman and Crane, 2011; Mihalcea and Nastase, 2012; Cook et al., 2014; Frermann and Lapata, 2016; Schlechtweg and Walde, 2020) estimate a probability distribution of words over their different senses, i.e., topics and (3) Clustering models (Mitra et al., 2015; Tahmasebi and Risse, 2017).

1.2 The DIACR-Ita task

The goal of the DIACR-Ita task (Basile et al., 2020a; Basile et al., 2020b) is to establish if a set of Italian words (target words) change their meaning from time period t_1 to time period t_2 (i.e., bi-

nary classification task). The organizers provide corresponding corpora C_1 and C_2 and a list of target words. Only these inputs may be used to train systems, which judge for each target word, whether it is changed or not. The task is the same as the binary sub-task of the SemEval-2020 Task 1 (Schlechtweg et al., 2020) competition.

2 Data

The DIACR-Ita data consists of many randomly ordered text samples that have no relationship to each other. Most of the text samples are complete sentences, but some are sentence fragments.

The 'early' corpus, C_1 has about 2.4 million text samples and 52 million tokens; the 'later' corpus, C_2 has about 7.8 million text samples and 738 million tokens. Each token is given in the corpora with its part-of-speech tag and lemma. The target word list consists of 18 lemmas. The POS and lemmas of the corpora are generated with the UDPipe (Straka, 2018) model ISDT-UD v2.5, which has an error rate of about 2%.

3 System Description

3.1 Overview

Because language is evolving, expressions, words, and sentence constructions in two corpora from different time periods about the same topic will be written in languages that are quite similar but slightly different. They will share the majority of their words, grammar, and syntax. We can observe a similar situation in languages from the same family, such as *Italian-Spanish* in Romance languages or *Czech-Slovak* in Slavic languages. These pairs of languages share a lot of common words, expressions and syntax. For some pairs, native speakers can understand and sometimes even actively communicate through a (low) language barrier.

Our system follows the approach from (Pražák et al., 2020)[1]. The main idea behind our solution is that we treat each pair of corpora C_1 and C_2 as different languages L_1 and L_2 even though the text from both corpora is written in Italian. We believe that these two languages L_1 and L_2 will be extremely similar in all aspects, including semantic. We train a separate semantic space for each corpus, and subsequently, we map these two spaces into one common cross-lingual space. We

use methods for cross-lingual mapping (Brychcín et al., 2019; Artetxe et al., 2016; Artetxe et al., 2017; Artetxe et al., 2018a; Artetxe et al., 2018b) and thanks to the large similarity between L_1 and L_2 the quality of transformation should be high. We compute cosine similarity of the transformed word vectors to classify whether the target words changed their sense.

3.2 Semantic Space Transformation

First, we train two semantic spaces from corpus C_1 and C_2. We represent the semantic spaces by a matrix $\mathbf{X}^s$ (i.e., a source space s) and a matrix $\mathbf{X}^t$ (i.e., a target space t)[2] using word2vec Skip-gram with negative sampling (Mikolov et al., 2013). We perform a cross-lingual mapping of the two vector spaces, getting two matrices $\hat{\mathbf{X}}^s$ and $\hat{\mathbf{X}}^t$ projected into a shared space. We select two methods for the cross-lingual mapping *Canonical Correlation Analysis (CCA)* using the implementation from (Brychcín et al., 2019) and a modification of the *Orthogonal Transformation* from *VecMap* (Artetxe et al., 2018b). Both of these methods are linear transformations. The transformations can be written as follows:

$$\hat{\mathbf{X}}^s = \mathbf{W}^{s \to t} \mathbf{X}^s \quad (1)$$

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

Finally, in all transformation methods, for each word w_i from the set of target words T, we select its corresponding vectors $\mathbf{v}^s_{w_i}$ and $\mathbf{v}^t_{w_i}$ from matrices $\hat{\mathbf{X}}^s$ and $\hat{\mathbf{X}}^t$, respectively ($\mathbf{v}^s_{w_i} \in \hat{\mathbf{X}}^s$ and $\mathbf{v}^t_{w_i} \in \hat{\mathbf{X}}^t$), and we compute cosine similarity between these two vectors. The cosine similarity is then used to generate a final classification output using different strategies, see Section 3.5 and 3.6.

3.3 Canonical Correlation Analysis

Generally, the CCA transformation transforms both spaces $\mathbf{X}^s$ and $\mathbf{X}^t$ into a third shared space o (where $\mathbf{X}^s \neq \hat{\mathbf{X}}^s$ and $\mathbf{X}^t \neq \hat{\mathbf{X}}^t$). Thus, CCA computes two transformation matrices $\mathbf{W}^{s \to o}$ for the source space and $\mathbf{W}^{t \to o}$ for the target space. The transformation matrices are computed by

---

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

[2] The source space $\mathbf{X}^s$ is created from the corpus C_1 and the target space $\hat{\mathbf{X}}^t$ is created from the corpus C_2.

minimizing the negative correlation between the vectors $\mathbf{x}_i^s \in \mathbf{X}^s$ and $\mathbf{x}_i^t \in \mathbf{X}^t$ that are projected into the shared space o. The negative correlation is defined as follows:

$$\operatorname*{argmin}_{\mathbf{W}^{s\to o},\mathbf{W}^{t\to o}} - \sum_{i=1}^{n} \rho(\mathbf{W}^{s\to o}\mathbf{x}_i^s, \mathbf{W}^{t\to o}\mathbf{x}_i^t) =$$

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

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

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

The submissions that use CCA are referred to as **cca-bin** and **cca-ranking** in Table 1. The **-bin** and **-ranking** parts refer to a strategy used for the final classification decision, see Section 3.5 and 3.6.

3.4 Orthogonal Transformation

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

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

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

3.5 Binary Strategy

We use different strategies for the binary classification output, but all have in common that they use continuous scores. The continuous score for each target word is computed as the cosine similarity between the two vectors from the earlier and later corpus.

In the case of the *binary strategy*, we assume a threshold t for which the target words with a continuous score greater than t changed meaning and words with the score lower than t did not. We know that this assumption is generally wrong (because using the threshold, we introduce some error into the classification), but we still believe it holds for most cases and it is the best choice. To estimate the threshold t, we used an approach called *binary-threshold* (**cca-bin** and **ort-bin** in Table 1). For each target word w_i we compute cosine similarity of its vectors $\mathbf{v}_{w_i}^s$ and $\mathbf{v}_{w_i}^t$, then we average these similarities for all words. The resulting averaged[3] value is used as the threshold.

3.6 Ranking Strategy

The *ranking strategy* is the second approach for generating a classification output (the submission result **cca-ranking** in Table 1). It uses the mean rank of repeated runs of each embedding pair. For each run, the target words are scored with a cosine distance. Then the distances for each embedding pair are sorted and a rank-order is assigned to each target. The rank-orders are averaged, to get a mean rank (and a standard deviation) for each target for each pair. Finally, ranks for all embedding pairs are averaged. The composite rank is used, along with an estimate of the associated cosine distance and its corresponding angle, to divide the target list into changed and unchanged sets. This does not work well; there are competing gaps in rank and distance estimates.

We use the number of embeddings, and not the total number of runs, to compute the standard error of the mean (which is standard deviation divided by the square root of samples).

4 Experimental Setup

To obtain the semantic spaces, we employ Skipgram with negative sampling (Mikolov et al., 2013). For the final submission, we trained the semantic spaces with 100 (the **ort-bin** submission)

---

[3]The **ort-bin** submission sets the threshold to be in the largest gap between the similarity values

and 150 (the **cca-bin** submission) dimensions for five iterations with five negative samples and window size set to five. Each word has to appear at least five times in the corpus to be used in the training. To train the semantic space, we used the lemmatized corpora. The dimensions 100 and 150 are selected based on our previous experiences with these methods (Pražák et al., 2020). Since we were able to submit four different submissions, we did not use the same dimension for both methods.

The **cca-ranking** submission uses the same settings and dimensions 100-105, 110-115, etc. up to 210-215, resulting in 72 different dimension sizes. It combines 40 runs on each of 72 embedding pairs, a total of 2880 runs.

For the **cca-bin** submission, we build the translation dictionary for the transformation of the two spaces by removing the target words from the intersection of their vocabularies. In the case of the **cca-ranking** submission, the dictionary in each run consists of up to 5000 randomly chosen common words for each semantic space.

The **random** submission represents output that was generated completely randomly.

4.1 Corpus variants

The organizers provided the corpora already tokenized in four different versions: original tokens; lemmatized tokens; original tokens with POS tag; lemmatized tokens with POS tag. We experimented with each of these variants, although in the end, we used results based only on lemmas. Figure
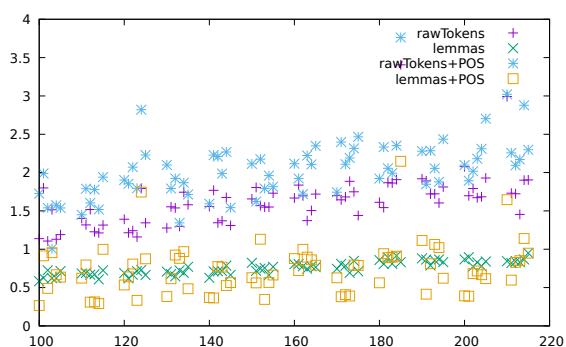

Figure 1: Standard deviation (of rank) versus embedding size for four versions of the corpora.

1 shows the mean standard deviation of rank for target words over forty runs for each of 72 different embedding sizes. The most consistent variant is the *lemmas* only.

5 Results

We submitted four different submissions. The accuracy results for each submission are shown in Table 1. The **ort-bin** system achieved the best accuracy of 0.944 and ranked first[4] among eight other teams in the shared task, classifying 17 out of 18 target words correctly. The **cca-bin** system achieved an accuracy of 0.889 (16 correct classifications out of 18). After releasing the gold labels, we performed an additional experiment with the **cca-bin** system achieving also an accuracy of 0.944 when the same word embeddings (with embeddings dimension 100 instead of 150) are used as for the **ort-bin** system. We found an optimal threshold for both systems, which makes them classify all the words correctly[5].

We believe that the key factor of the success of our system is the sufficient size of the provided corpora. Thanks to that, we were able to train semantic spaces of good quality and thus achieve good results.

System	Accuracy
cca-bin	.889
ort-bin	.944
cca-ranking	.778
random	.500

Table 1: Results for our final submissions.

6 Conclusion

Our systems based on Canonical Correlation Analysis and Orthogonal Transformation achieved the best accuracy of 0.944 in the shared task and ranked first among eight other teams. We showed that our approach is a suitable solution for the *Lexical Semantic Change* detection task. Applying a threshold to semantic distance is a sensible architecture for detecting the binary semantic change in target words between two corpora. Our *binary-threshold* strategy succeeded quite well.

This task provided plenty of text to build good word embeddings. Corpora with much smaller amounts of data might have increased the random variation between the earlier and later embeddings, which would have given our method problems. A flaw in our technique is that semantic vec-

---

[4]We share the first place with another team that achieved the same accuracy.

[5]That is, 100% accuracy was possible with the continuous scores of both methods if we only had an oracle to set the threshold.

tors are based on all senses of a word in the corpus. We do not yet have tools to tease out what kinds of changes are implied by a particular semantic distance between vectors. We considered using the part of speech data in the corpora since different parts of speech for the same lemma are likely different senses. But placing the POS in the token, like using inflections instead of lemmas, results in many more, less well-trained semantic vectors, as suggested by Figure 1.

Acknowledgements

References

[Artetxe et al.2016] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2016. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2289–2294, Austin, Texas, November. Association for Computational Linguistics.

[Artetxe et al.2017] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 451–462, Vancouver, Canada, July. Association for Computational Linguistics.

[Artetxe et al.2018a] Mikel Artetxe, Gorka Labaka, , and Eneko Agirre. 2018a. Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pages 5012–5019.

[Artetxe et al.2018b] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018b. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 789–798, Melbourne, Australia, July. Association for Computational Linguistics.

[Bamman and Crane2011] David Bamman and Gregory Crane. 2011. Measuring historical word sense variation. In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, JCDL '11, page 1–10, New York, NY, USA. Association for Computing Machinery.

[Basile et al.2020a] Pierpaolo Basile, Annalina Caputo, Tommaso Caselli, Pierluigi Cassotti, and Rossella Varvara. 2020a. DIACR-Ita @ EVALITA2020: Overview of the EVALITA2020 Diachronic Lexical Semantics (DIACR-Ita) Task. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Online. CEUR.org.

[Basile et al.2020b] Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro. 2020b. Evalita 2020: Overview of the 7th evaluation campaign of natural language processing and speech tools for italian. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.

[Brychcín et al.2019] Tomáš Brychcín, Stephen Taylor, and Lukáš Svoboda. 2019. Cross-lingual word analogies using linear transformations between semantic spaces. *Expert Systems with Applications*, 135:287–295.

[Cook et al.2014] Paul Cook, Jey Han Lau, Diana McCarthy, and Timothy Baldwin. 2014. Novel word-sense identification. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1624–1635, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.

[Eger and Mehler2016] Steffen Eger and Alexander Mehler. 2016. On the linearity of semantic change: Investigating meaning variation via dynamic graph models. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 52–58, Berlin, Germany, August. Association for Computational Linguistics.

[Frermann and Lapata2016] Lea Frermann and Mirella Lapata. 2016. A Bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics*, 4:31–45.

[Gulordava and Baroni2011] Kristina Gulordava and Marco Baroni. 2011. A distributional similarity

approach to the detection of semantic change in the Google books ngram corpus. In *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 67–71, Edinburgh, UK, July. Association for Computational Linguistics.

[Hamilton et al.2016a] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016a. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2116–2121, Austin, Texas, November. Association for Computational Linguistics.

[Hamilton et al.2016b] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016b. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany, August. Association for Computational Linguistics.

[Kutuzov et al.2018] Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1384–1397, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

[Mihalcea and Nastase2012] Rada Mihalcea and Vivi Nastase. 2012. Word epoch disambiguation: Finding how words change over time. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 259–263, Jeju Island, Korea, July. Association for Computational Linguistics.

[Mikolov et al.2013] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *Proceedings of workshop at ICLR*. arXiv1301.3781.

[Mitra et al.2015] Sunny Mitra, Ritwik Mitra, Suman Kalyan Maity, Martin Riedl, Chris Biemann, Pawan Goyal, and Animesh Mukherjee. 2015. An automatic approach to identify word sense changes in text media across timescales. *Natural Language Engineering*, 21(5):773–798.

[Pražák et al.2020] Ondřej Pražák, Pavel Přibáň, Stephen Taylor, and Jakub Sido. 2020. Uwb at semeval-2020 task 1: Lexical semantic change detection. In *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020)*, Barcelona, Spain, Sep. Association for Computational Linguistics.

[Rosenfeld and Erk2018] Alex Rosenfeld and Katrin Erk. 2018. Deep neural models of semantic shift. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,*

Volume 1 (Long Papers), pages 474–484, New Orleans, Louisiana, June. Association for Computational Linguistics.

[Schlechtweg and Walde2020] Dominik Schlechtweg and Sabine Schulte im Walde. 2020. Simulating lexical semantic change from sense-annotated data. In A. Ravignani, C. Barbieri, M. Martins, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, K. Mudd, and T. Verhoef, editors, *The Evolution of Language: Proceedings of the 13th International Conference (EvoLang13)*.

[Schlechtweg et al.2017] Dominik Schlechtweg, Stefanie Eckmann, Enrico Santus, Sabine Schulte im Walde, and Daniel Hole. 2017. German in flux: Detecting metaphoric change via word entropy. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 354–367, Vancouver, Canada, August. Association for Computational Linguistics.

[Schlechtweg et al.2018] Dominik Schlechtweg, Sabine Schulte im Wlade, and Stefanie Eckmann. 2018. Diachronic usage relatedness (durel): A framework for the annotation of lexical semantic change. In *Proceedings of NAACL-HLT 2018*, pages 169–174.

[Schlechtweg et al.2019] Dominik Schlechtweg, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. 2019. A wind of change: Detecting and evaluating lexical semantic change across times and domains. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 732–746, Florence, Italy, July. Association for Computational Linguistics.

[Schlechtweg et al.2020] Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. SemEval 2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020)*, Barcelona, Spain, Sep. Association for Computational Linguistics.

[Straka2018] Milan Straka. 2018. UDPipe 2.0 prototype at CoNLL 2018 UD shared task. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 197–207, Brussels, Belgium, October. Association for Computational Linguistics.

[Tahmasebi and Risse2017] Nina Tahmasebi and Thomas Risse. 2017. Finding individual word sense changes and their delay in appearance. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 741–749, Varna, Bulgaria, September. INCOMA Ltd.

[Tahmasebi et al.2018] Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2018. Survey of computational approaches to lexical semantic change. *arXiv preprint arXiv:1811.06278*.