TU DORTMUND

ADVANCED TEXT MINING METHODS

SEMINAR

Diachronic embeddings

Exploring Word2Vec with alignment methods for Contextual Event Tracking: A Research Report

Lecturers:

Prof. Dr. Carsten Jentsch

M. Sc. Kai-Robin Lange

Author: Imene KOLLI

March 27, 2023

Contents

1 Introduction

Language constantly evolves throughout time since it is a reflection of culture and society. For instance, speakers might shift their language usage and word choice in response to important occurrences like wars, political movements, or technical developments (Kutuzov et al., 2017). We may track the progression of events and learn how they were perceived and interpreted in various historical contexts by examining the semantic shift in documents from various eras (Bamler and Mandt, 2017). We may also spot changes in attitudes, beliefs, and values that went along with these events by analyzing the language used in news articles, literature, and other written materials from past periods (Hamilton et al., 2016). These revelations can aid in our understanding of the events' historical and cultural background as well as the effects they had on society.

Our purpose in this project is to study the changes in meaning of words and how those changes reflect broader cultural and social changes over time. It involves the analysis of the semantic evolution of words distributionally across many historical periods, using diachronic embeddings methods.

In this study, we focus on the semantic evolution of polysemous words and their contextual use in different time periods. Our approach seeks to showcase the use of Word2Vec to extract diachronic word embeddings. Our findings demonstrate the effectiveness of this approach in detecting changes in word meaning and revealing the historical events that triggered those changes.

In section **4.1**, the study focuses on extracting the diachronic embeddings of target words to determine the time points in which semantic shifts had occurred. The study employs the cosine similarity measure along with the PELT algorithm to estimate the change-points in the self-similarity time series of the target words.

In section **4.2**, we visualize the trajectory of target words in their contextual semantic associations to gain insights into the real-world events that triggered the changes in their meaning and usage. The study also utilizes context word embeddings to extract attributes about these events, such as their geographic locations and the groups of people impacted by them. By analyzing the semantic associations captured by the diachronic embeddings, this study provides a means to track and comprehend real-life events, offering a deeper understanding of the social and cultural implications of the semantic shifts. In section **4.3**, we use the time series K-means algorithm with dynamic time wrapping and Barycenter averaging methods to cluster embeddings that evolve in similar patterns to understand the complex interplay between language use and real-life events.

Finally, Section **5** summarizes the key findings of the research and draws conclusions about the effectiveness of diachronic embeddings, generated by Word2Vec, in tracking events and identifying semantic shifts in language use over time.

2 Problem statement

2.1 Data Assessment

Our data analysis will center on a dataset of news articles sourced from The New York Times. The corpora comprises over 2 million articles published between 1980 and 2018, spanning a broad spectrum of topics such as politics, economics, and social issues. To ensure a comprehensive representation of language use during this period, we divided the articles into 22 distinct corpora, each representing a different time period. Each corpus consists of 90,000 articles published on the same year, all of which contain at least one of our 3 selected target polysemous words. This approach allows us to focus on the semantic evolution of these words over time and gain insights into the cultural changes that have occurred. Overall, this dataset provides a rich and diverse resource for diachronic embeddings research.

In this study, we choose: "abuse," "war", and "market" as our target words. We justify the selection of these words based on their relevance to a range of social, political, and economic issues that have evolved over time. Table 1 contains a more in depth explanation to the choice of these words.

Target word	word relevance
Abuse	This word is relevant to a broad range of social and political issues, including domestic violence, substance abuse, and political corruption. (abu, 2022)
War	This word is relevant to understanding the cultural and social changes that have occurred over time in relation to conflict and military engagements. (war, 2023)
Market	This word is relevant to understanding economic trends and the evolution of business practices over time. (mar, 2023)

Table 1: Target words with explanation to their relevance in a diachronic analysis

Overall, the choice of these three target words highlights their relevance to various contemporary social, political, and economic problems.

2.2 Project Objectives

The objective of this study is to track the evolution of events through time using diachronic word embeddings generated by the Word2Vec algorithm. Our first step involves tokenizing sentences of each corpus, ensuring that each sentence contains at least one of our target words. Following this, we preprocess the text by eliminating stop words, punctuation, and performing lemmatization on the words.

After preparing the data, our next objective is to utilize the Word2Vec neural network to generate embeddings that represent our words as vectors in a multi-dimensional space. We rely on the Orthogonal Procrustes method to train and align multiple Word2Vec models. Through this process, diachronic embeddings are generated that capture the changing meaning and usage of words over time.

Once the diachronic embeddings are generated, we seek to utilize time series methods to detect semantic shifts that have occurred in our target words. Specifically, we analyze changes in distributional patterns over time and model them using PELT algorithm and time series k-means, to identify shifts in semantic associations.

Finally, we aim to contextualize the results of the analysis within the cultural and social context of the time period. This involves considering how changes in language use reflect changes in attitudes, beliefs, and values, and how these changes relate to broader social and cultural trends.

3 Methods

The present section aims to give a summary of the statistical techniques employed in the study.

3.1 Data Preprocessing

To analyze each corpus (C_t), we extract only the sentences that contain at least one of our target words. These sentences are then preprocessed by removing irrelevant words such as stop words, punctuation, and numbers. Additionally, the words are lemmatized to reduce the number of unique words that need to be analyzed and to ensure that similar words are treated as the same entity. The final outcome of this process is a separate corpus for each time period to be analyzed, where each corpus consists of cleaned sentences.

3.2 Diachronic embeddings: Time series of static word embeddings

Diachronic embeddings are a series of word embeddings, each of which captures the meaning of words in a certain time period. They are generated by training Word2Vec models on multiple corpora from different time periods. By training the models on each time period separately, the resulting embeddings for each word form a time series of vectors, representing the meaning of the word at each point in time. Using this approach, the meaning of a given word can be tracked over time and analyzed. For example, the diachronic embeddings of the word "gay" can reveal how its semantics and usage have evolved through time from meaning of "cheerful" to its more recent sense in association with homosexuality. (Hamilton et al., 2016)

The Skip-gram with negative sampling Word2Vec - SGNS

The process of generating diachronic embeddings for our target words involves training multiple Word2Vec models. In this study we use the SGNS algorithm. SGNS is a prediction-based model, trained to predict the probability of a word appearing in the context of other words in a sentence. This method is used to generate vector representations of all words in the corpus. The basic premise is that words that share similar contexts are likely to have similar meanings and, therefore, should have similar vector representations (Mikolov et al., 2013). Depending on the *window size* hyperparameter, the algorithm is trained to consider the context words for every word in the corpus. Specifically, the algorithm treats each word in the training corpus as a target word and tries to predict the words that appear within a fixed-size window around the target word. (Illustrated Word2Vec, 2019)

Below, Figure 1 shows the architecture of the SGNS neural network.

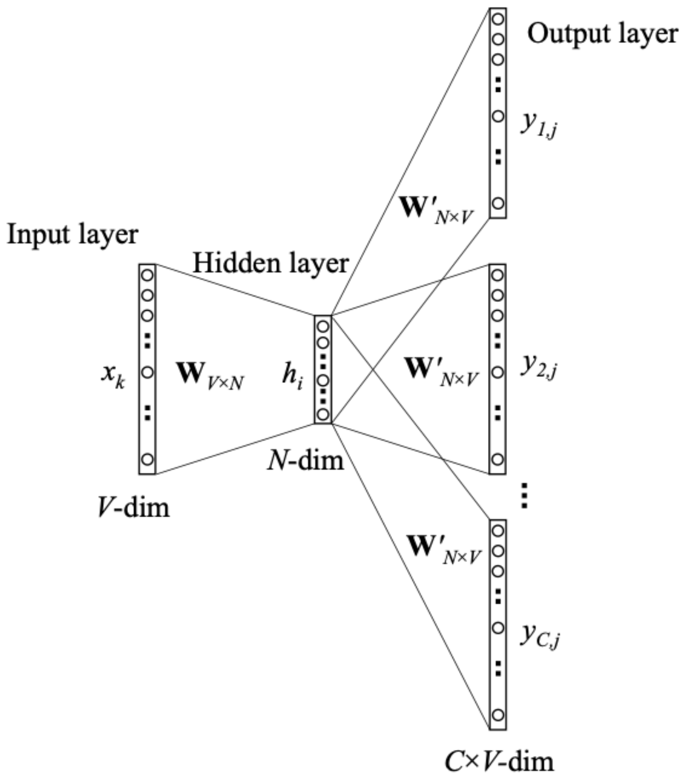

Figure 3: The skip-gram model.

Figure 1: The SGNS architecture, a model for obtaining word embeddings. Source: (Rong, 2014), fig. 3.

SGNS is particularly effective at identifying semantic relationships between words, such as synonyms, antonyms, and analogies. This is possible because each word is embedded as a vector in a high-dimensional space, where the distance between these embeddings reflects the similarity between the corresponding words (Hamilton et al., 2016). To capture different types of semantic relationships, the window size of the algorithm can be adjusted depending on the task at hand. A large window size, such as 15 or higher, is better suited for capturing global semantic relationships between words. In this case, a high degree of similarity between two word embeddings suggests that the words are related. This is particularly useful for analyzing the diachronic changes in the semantics of our three target words, where we intend to capture the changes in their related words set over time.(Illustrated Word2Vec, 2019)

3.3 Model alignment: Vector mapping to combine vector spaces

The training of multiple Word2Vec models generates multiple embeddings for each word in the vocabulary. However, vectors generates by one Word2Vec model are only comparable within the vector space of that model. This is due to the random intialization of the algorithm leading to numerically different vectors every training, while preserving the same pairwise similarities. To be able to compare embeddings generated by different models, the need to map these embeddings into one vector space arises. Model alignment is a technique used to map the vectors generated by multiple models to the vector space of one of those models. (Kutuzov et al.)

Orthogonal Procrustes

One approach to align Word2Vec models is to use Orthogonal Procrustes analysis. The first step is to extract the normalized unit vector embeddings of the common vocabulary by intersecting the vocabularies of each model. Then, the Procrustes problem 3.1 is solved by identifying the optimal rotation matrix $\hat{R}$ for each model's vectors to align with the chosen reference model's vector space. Hamilton et al. (2016)

$$\hat{R} := argmax_{R^T R=I} ||Y - XR||^2 = ||Y||^2 + ||XR||^2 - 2trace(Y^T XR) \tag{3.1}$$

where Y and X are the vectors in the reference space and the second model's space respectively. Finding the optimal rotation matrix reduces to maximizing 3.2.

$$trace(Y^T XR) = trace(UDV^T r) = trace(V^T RU.D) \tag{3.2}$$

where U, D, and V are the rotation, scaling, and reflection components of the singular value decomposition of the matrix product of X and Y. The problem is then solved by $\hat{R} = VU^T$. (Schönemann)

Finally, the product matrix of X and $\hat{R}$ produces aligned vectors of X to Y. (Hamilton et al., 2016)

3.4 Inference methods: tracking semantic associations

In order to infer semantic shift after extracting diachronic embeddings, we rely on different techniques such as the similarity measure to extract multiple types of insights about the semantics of our target words. We also apply time series clustering methods and change-point detection using the Pruned Exact Linear Time (PELT) algorithm.

The cosine similarity

One way to infer insights from diachronic embeddings, is to use the cosine similarity 3.3 between embeddings of the words across time (Hamilton et al., 2016). A self-similarity approach can be used to measure the change in the context of the target word between the past periods and the latest time period, which is kept as a reference point.

$$S_C(e_t^{w_i}, e_T^{w_i}) := \frac{e_t^{w_i} \cdot e_T^{w_i}}{||e_t^{w_i}|| \cdot ||e_T^{w_i}||} \tag{3.3}$$

where $e_t^{w_i}$ is the embedding of the word w_i in time period t and T is the latest period. By applying this approach, we obtain a time series of similarity values between the target word in past periods and its latest period's context. Furthermore, this method can be expanded to track the evolution of the target word by observing its most similar context words over time 3.4.

$$S_t^{w_i} = e_t^{w_j} : w_j \in N_n(w_i, t) \tag{3.4}$$

where $N_n(w_i, t)$ is the set of n most similar context words of w_i at time t. Extracting the set $S_t^{w_i}$ allows for a more detailed analysis of changes in the context of the target word. This enables us to gain insights into how the word was interpreted by speakers, as well as the occurrences that influenced the changes in context.

To track the trajectory of a target word between different semantic contexts over time, we also examine the set of most similar context word embeddings, extracted from the latest period, to every diachronic embedding of the target word. This allows us to visualize the changes in the word's meaning and usage over time, after reducing the dimensionality of the embeddings into 2-dimensional vectors. Hamilton et al. (2016)

PELT: Pruned Exact Linear Time

The PELT algorithm is a statistical technique that can be utilized to detect change-points in time-series data. Specifically, we can apply the algorithm to the self-similarity time series of the target word. The algorithm works by dividing the time series into segments and minimizing the sum of the within-segment variances. The algorithm is regularised with a penalty on the number of segments by the cost parameters. (Killick et al.)

The algorithm starts by computing the cost $C(t, 1)$ of the entire time series being in one segment. 3.5.

$$C(t, 1) = Var(S_C(e_{t=1}^{w_i}, e_T^{w_i}), ..., S_C(e_{T-1}^{w_i}, e_T^{w_i}), 1) \tag{3.5}$$

The PELT algorithm iteratively increases the number of partitions k while computing the cost in equation 3.6. The algorithm continues to increase k until it identifies the optimal change-points that minimize the total cost. (Killick et al.)

$$C(t,k) = min\{C(j,k-1) + Var(S_C(e_{j+1}^{w_i}, e_T^{w_i}), ..., 1)\}_{j\in\{1,...,t-1\}} \quad (3.6)$$

Time-series K-Means with Dynamic Time Wrapping

Time-series K-Means with Dynamic Time Warping is an unsupervised technique that can be used to cluster diachronic embeddings based on patterns in their trends. It involves applying the K-Means algorithm to the time series data of the embeddings, while using the Dynamic Time Warping (DTW) algorithm as a distance measure to account for differences in the time series shapes and lengths. Dynamic time wrapping is a technique that aligns two time series of lengths N and M, by warping one of them to better match the other using a path P, allowing for differences in their shapes and patterns. This technique is particularly useful in diachronic embeddings, where the time series data may exhibit variations in frequency, amplitude, or other features over time. (Gold and Sharir, 2018)

$$DTW(x,y) = \min_P \sqrt{\sum_{(i,j)\in P} d(x_i, y_j)^2} \quad (3.7)$$

where d is the euclidean distance and $P = [p_1, ..., p_L]$ is the wrapping path of length L, with p_l satisfying multiple properties (Müller, 2007, Section 3.2.1):

- The boundary property: $p_1 = (1,1)$ and $p_L = (N,M)$.
- The monotonicity property: $p_l = (n_l, m_l)$ where $n_1 < n_l < n_L$ and $m_1 < m_l < m_L$ $\forall l : 1 < l < L$
- The step size condition: $p_{l-1} - p_l \in \sum := \{(1,0),(0,1),(1,1)\}$ for $l \in [1:L]$

The time-series K-Means works by iteratively assigning each data point to the closest cluster center and then updating the cluster centroids to minimize the sum of the squared DTW distances to their assigned points (Dau et al., 2018). The centroids are updated by computing the mean point of each cluster, using the Barycenter averaging method (Petitjean et al., 2011). Using the DTW algorithm, the embeddings time series of each word is wrapped with the barycenters to assign the word into its closest cluster.

4 Applications

In this project, `Python 3.10.9` is used for statistical computations and graphical visualizations. The following packages are used:

- `plotly, Version 5.13.1` (Inc., 2015)
- `gensim, Version 4.3.1` (Rehurek and Sojka, 2011)
- `statsmodels, Version 0.12.2` (Seabold and Perktold, 2010)

In the following sections, we use the introduced methods to track semantic shifts using diachronic embeddings.

4.1 Change-points and self-similarity

Analyzing change-points and self-similarity in diachronic embeddings is crucial for gaining insights into the timing and nature of semantic shifts. To begin this analysis, we examine the self-similarity of our target words. To ensure interpretability, we select the latest period $t = 2018$ as a reference point since it is the closest period to the present, assuming that we all have a common understanding of the meanings of the target words in the present time. As shown in Figure 2, words can undergo significant changes in their contextual meaning over time, which may not align with our current interpretation of the word.

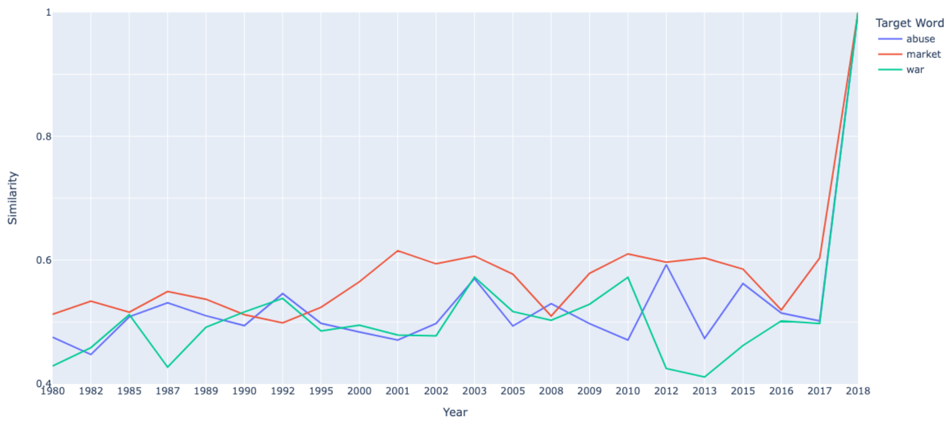

Figure 2: Graph shows the self-similarity time series of the target words: war, abuse, market.

To determine the timing of semantic shifts in our three target words, we apply the PELT algorithm to their self-similarity time series. The algorithm partitions the time series into segments that correspond to different semantic contexts, and identifies the time-points where significant changes in context occur. Figure 5 (see appendix) displays the different partitioning time-points identified by the algorithm. These findings can be used in subsequent analysis to explore the relationship between the identified time-points and the historical events that influenced the changes in meaning. For example, we can examine news reports from that time period to understand the events that the speakers experienced and how they may have influenced the semantic shifts of the target words.

4.2 Word trajectory and context words

The generated diachronic embeddings are also used to follow the trajectory of our target words through time. We start by extracting the embeddings of the most similar context words to the target words in $t = 2018$, and the time series of embeddings of each target. For visualization purposes, we reduce the dimentionality of all embeddings to 2 dimension using the t-distributed stochastic neighbour embedding (t-SNE) technique (van der Maaten and Hinton, 2008). In Figure 3, we seek to follow the trajectory of the target *"abuse"* over past decades starting from $t = 1980$. The graph shows that in the early periods, the word's contextual associations were more aligned with narcotics as in the excessive use of substances like drugs and alcohol. Through time, the word shifted its semantics to include contexts more related to violence and sexual abuse, particularly against children and women. It's interesting to notice the sense narrowing that occurred on the word, where in the most recent periods, the word usage mostly associated with child abuse semantics. That is why context words that show relatedness, such as bishop and church, appear in the neighbourhood of the word.

The presence of related context words, like *"larry"*, highlights the effectiveness of diachronic embeddings in tracking events by leveraging the use of context words. In this context, *"larry"* may refer to Pope Francis's visit to Chile in 2018, demonstrating how diachronic embeddings can capture the semantic connections between words and events over time[1].

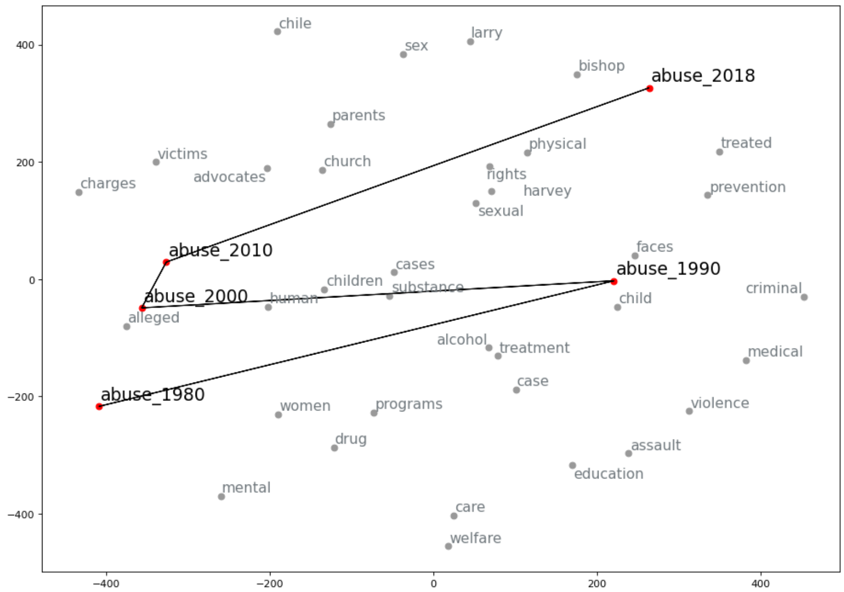

Figure 3: Graph shows the trajectory of the word *"abuse"* over past decades from 1980 to 2018.

[1]https://www.theguardian.com/world/2018/may/02/pope-francis-chile-sexual-abuse-scandal-part-of-problem

Another analysis on the frequency of reported abuse types that occurred over the past periods is conducted using the generated diachronic embeddings. The aim is to determine which types of abuse occurred during different periods by tracking the similarity between a fixed set of context word embeddings and the target embeddings over time. The choice of the context words is made to reflect the types of abuse we are interested in exploring. Figure 4 visualizes the relations between multiple context words and *"abuse"*. It's interesting to notice the similarity time series of the word "Financial," which remains consistently low between $t = 1987$ and $t = 2001$. This pattern coincides with the U.S.'s transition from relatively low inflation rates to virtual price stability during this period[2]. As a result, the semantic association of the word *"abuse"* moved further away from the concept of finance, resulting in low similarity values.

The plot results are also consistent with our previous findings, highlighting the narrowing of the semantic associations of the target and the shift in the meaning of "abuse" towards violent sexual acts committed against women and children.

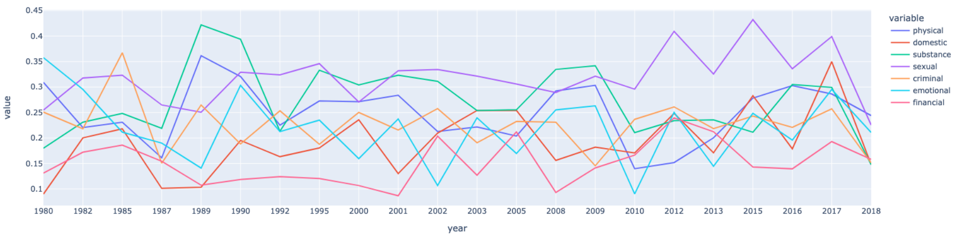

Figure 4: Graph shows the similarity between multiple context words and *"abuse"* from 1980 to 2018.

While some words, such as *war*, may not undergo significant linguistic shifts, the context in which they are used can change, leading to the acquisition of new semantic associations. For instance, Figure 6 (see appendix) shows how the word *war* can undergo a semantic shift and its embedding vector can be influenced by its association with the geographic location of the war, resulting in fluctuations in the self-similarity time series.

4.3 Diachronic embeddings clustering

Lastly, we apply the time series K-means algorithm to cluster diachronic embeddings that exhibit similar patterns of change. To investigate the semantic trends of our target words, we choose two context words for each target and test the algorithm to group them into their respective clusters. The clustering results of our approach on nine selected words are shown in Figure 7 (see appendix), and summarized in Table 2 below.

[2]https://papers.ssrn.com/sol3/papers.cfm?abstract$_i d = 2183357$

Table 2: Summary of clustering results after applying time series k-means with dynamic time wrapping and barycenter averaging methods on diachronic embeddings.

Chosen Word	Cluster
War	1
Syria	1
Iraq	1
Market	2
Financial	2
Crisis	1*
Abuse	3
Drug	3
Alcohol	3

Overall, the algorithm effectively grouped the target words and their respective context words into their expected clusters. However, the word *"Crisis"* did not cluster as anticipated. Despite the occurrence of multiple financial crises in the past periods, it was instead grouped into cluster 1, which is more closely associated with wars. This result can be attributed to the fact that there were more wars than financial crises during the selected time periods, which may have influenced the clustering of the word.

5 Summary

Diachronic embeddings have proven to be an effective tool for tracking events and identifying semantic shifts in contextual associations of target words. One of the advantages of diachronic embeddings is the ability to track and capture real-life events through the use of Word2Vec, alignment, and similarity measures. This allowed us to extract valuable information about the events that triggered the semantic shifts and the impact they had on language use. Inference methods like following a word's trajectory were used to extract information about these events like the impacted groups and other related attributes, offering a deeper understanding of the social and cultural implications of the semantic shifts. The use of diachronic embeddings in natural language processing, historical linguistics, and social sciences can help researchers to better understand the complex interplay between language use and real-life events, and to track the evolution of language use and meaning over time. However, diachronic embeddings are still currently limited in their ability to capture more nuanced changes or account for variations in language use across different social groups or regions. That is because they may overlook subtle changes in meaning that are not reflected in the training data.

Bibliography

abuse. n,. In *OED Online*. Oxford University Press, 12 2022. URL https://www.oed.com/view/Entry/821?rskey=zdv74Cresult=1eid.

market. n,. In *OED Online*. Oxford University Press, 03 2023. URL https://www.oed.com/view/Entry/114178?rskey=uj3iRQresult=1eid.

war. n,. In *OED Online*. Oxford University Press, 03 2023. URL https://www.oed.com/view/Entry/225589?rskey=gBQTQdresult=1isAdvanced=falseeid.

Robert Bamler and Stephan Mandt. Dynamic word embeddings. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 380–389. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/bamler17a.html.

Hoang Anh Dau, Eamonn Keogh, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, Yanping, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, Gustavo Batista, and Hexagon-ML. The ucr time series classification archive, October 2018. https://www.cs.ucr.edu/ eamonn/time$_{s}eries_{d}ata_{2}$018/.

Omer Gold and Micha Sharir. Dynamic time warping and geometric edit distance: Breaking the quadratic barrier. *ACM Trans. Algorithms*, 14(4), aug 2018. ISSN 1549-6325. doi: 10.1145/3230734. URL https://doi.org/10.1145/3230734.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 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 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1141. URL https://aclanthology.org/P16-1141.

Illustrated Word2Vec. illustrated-word2vec, 2019. URL https://jalammar.github.io/illustrated-word2vec/. [Online; accessed 26. Mar. 2023].

Plotly Technologies Inc. Collaborative data science, 2015. URL https://plot.ly.

R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. URL https://arxiv.org/abs/1101.1438v3.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. Diachronic word embeddings and semantic shifts: a survey. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1384–1397. Association for Computational Linguistics. URL `https://aclanthology.org/C18-1117`.

Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. Tracing armed conflicts with diachronic word embedding models. In *Proceedings of the Events and Stories in the News Workshop*, pages 31–36, Vancouver, Canada, August 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-2705. URL `https://aclanthology.org/W17-2705`.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

Meinard Müller. *Information Retrieval for Music and Motion*. Springer Verlag, 2007. ISBN 3540740473.

François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, 2011. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2010.09.013. URL `https://www.sciencedirect.com/science/article/pii/S003132031000453X`.

Radim Rehurek and Petr Sojka. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2), 2011.

Xin Rong. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*, 2014.

Peter H. Schönemann. A generalized solution of the orthogonal procrustes problem. 31(1):1–10. ISSN 1860-0980. doi: 10.1007/BF02289451. URL `https://doi.org/10.1007/BF02289451`.

Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL `http://jmlr.org/papers/v9/vandermaaten08a.html`.

Appendix

A Additional figures

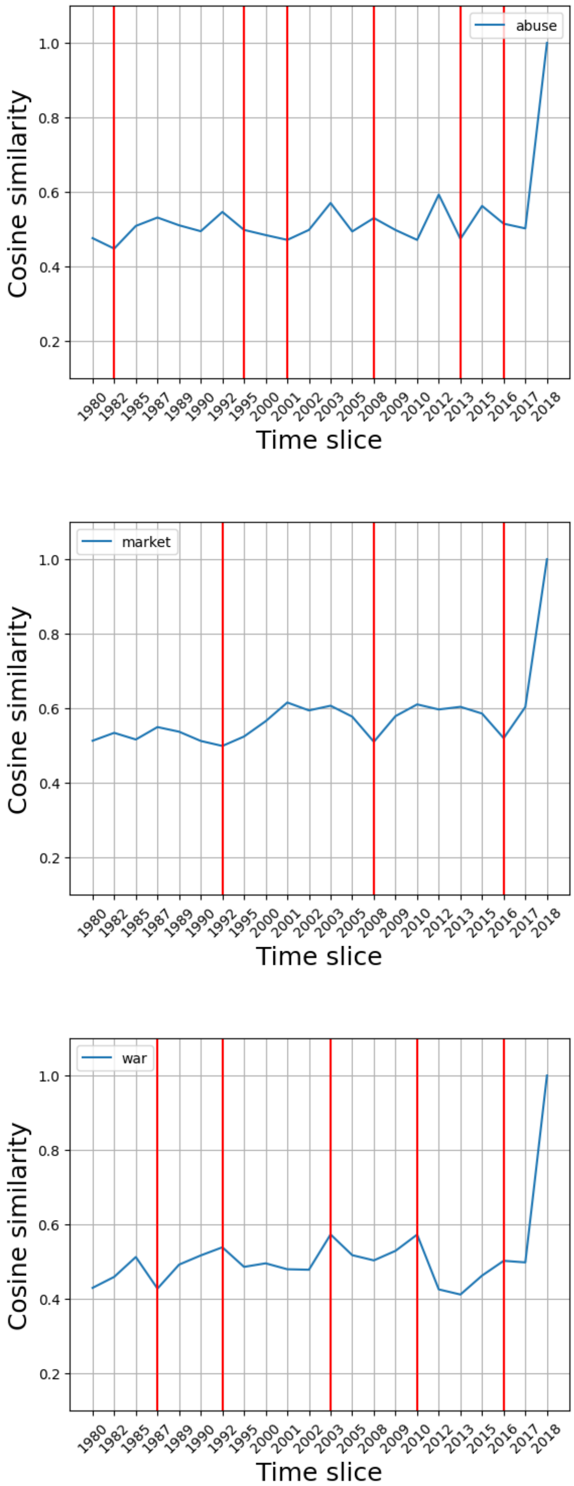

Figure 5: Graph shows the identified change-points in the words' self-similarity time series by the PELT algorithm

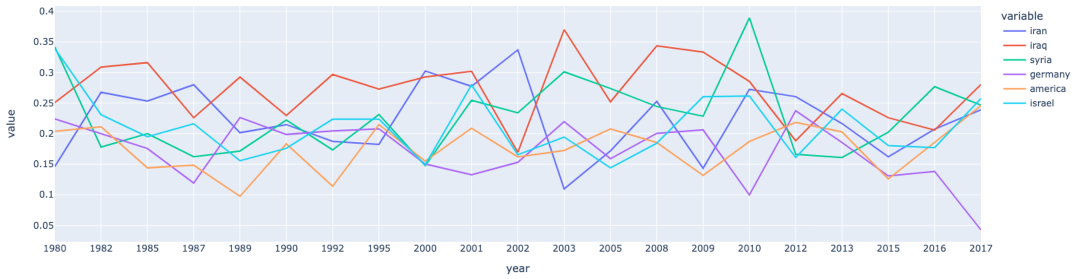

Figure 6: Graph shows the similarity between multiple context words and *"war"* from 1980 to 2018.

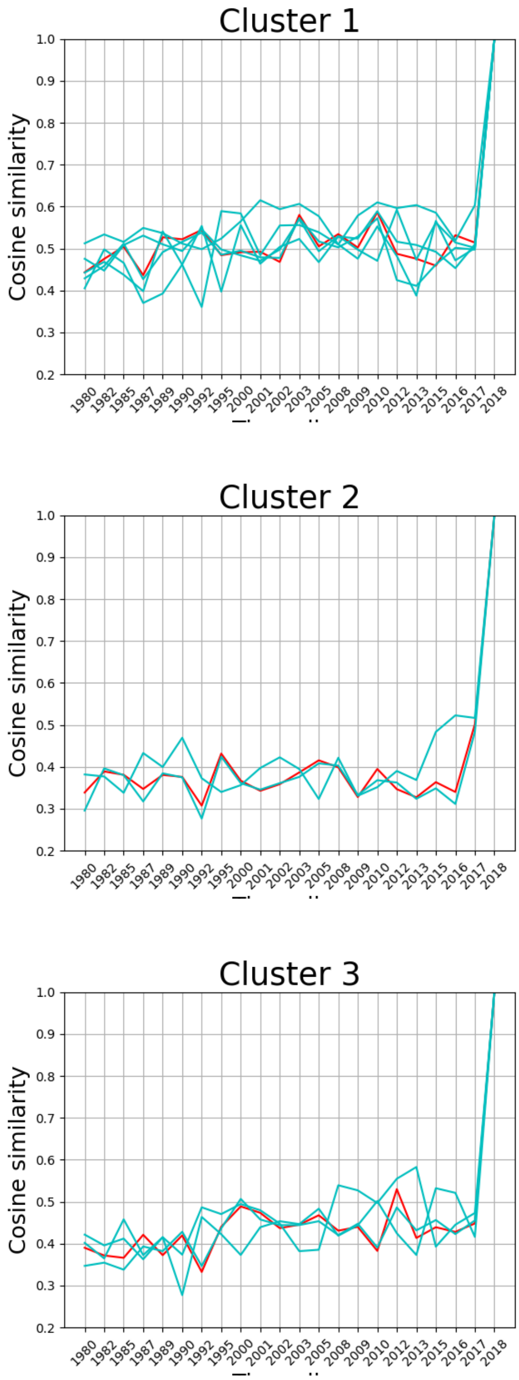

Figure 7: Graph shows the Time-series K-Means clustering of 9 words' self-similarity time series into 3 clusters. The red time series represents the barycenters of the clusters.