Analyzing Semantic Shift Across Time in Two Languages

Hugo Colinet
Department of Computer Science
Technische Universität Darmstadt
Darmstadt, Germany
hugo.colinet
@stud.tu-darmstadt.de

Tanbir Rahman
Department of Computer Science
Technische Universität Darmstadt
Darmstadt, Germany
aonik_r@yahoo.de

Katharina Anja Wallrabenstein
Department of Computer Science
Technische Universität Darmstadt
Darmstadt, Germany
katharina.wallrabenstein
@stud.tu-darmstadt.de

Siyang Zhang
Department of Computer Science
Technische Universität Darmstadt
Darmstadt, Germany
siyang.zhang
@stud.tu-darmstadt.de

Abstract

German and English are historically close languages that share common roots. Therefore, they have always comprised a significant number of semantically identical words. As time passes, words undergo semantic changes in both languages. Formerly distant words can semantically converge towards each other; other words that used to be close can diverge from one another or undergo a parallel change, acquiring a new common meaning. In this work, we develop a process to identify these specific relative semantic changes and we test it on a dataset of time-specific English and German word embeddings resulting from the analysis of texts from the 19th and the 20th century.

1 Introduction

Human languages and their components are subject to constant change over time. Changes in the lexical meaning of a word, as opposed to changes in grammar or form, can be categorized as *semantic shift* or *semantic change* (Bloomfield, 1933).

Thus *semantic change* describes how a word changes its meaning throughout time: for example, the word "gay" used to mean *happy* or *cheerful* in the early 1900s, but shifted its meaning to *homosexual* in recent times (Hamilton et al., 2016).

This paper compares the semantic changes in two related languages, English and German. Both being part of the Germanic language family, they share a common ancestor in the Proto-Germanic language (Konig and Van der Auwera, 2013). This can still be perceived today in the two languages' many similarities. However, as the languages drifted apart throughout the centuries, so did the semantics of their vocabulary: in the 13th century, the English word "hound" and the German word "Hund" meant the same thing, namely a dog. While the German word retained its meaning until today, the English word underwent a semantic change. The word "hound" now refers to a specific type of dog that is bred for hunting. (Fortson IV, 2017).

While there have been studies employing machine learning to examine the shift of a single language over time (Giulianelli et al. (2020), Hamilton et al. (2016)) and some comparing different languages at the same point in time (Gong et al. (2020)), our approach is novel in that it compares two languages at two separate points in time. Using a dataset containing word embeddings calculated separately for each decade, we create a single embedding space consisting of German and English, split into two time periods: "old" data from 1800 to the 1899, and "new" embeddings from 1900 to 1999. The examined time frame ranges from the early 19th century to the late 20th century.

We analyze this vector space by comparing the cosine similarities of different word pairs in the "old" and "'new" data. We aim to find word pairs such as the aforementioned *"hound"* and *"Hund"* that used to be similar but have drifted apart or *diverged*. We also look for *converging* word pairs, whose new interpretations are closer to each other. Additionally, we try to find terms that have undergone *parallel* changes, meaning they evolved in a similar direction.

Contributions

- We examine English and German for the occurrence of semantic change across different languages using pre-trained SGNS (Word2Vec) embeddings and a cosine-similarity based approach.

- We successfully identify English-German word pairs that underwent semantic change in either a *parallel*, *diverging*, or *converging* or manner.

- Despite doing so, the number of word pairs showing semantic change is limited. We attribute this to embedding quality resulting from the data available to us. We hope that future work could address this limitation by incorporating larger and better quality corpora.

2 Related work

Semantic Change Laws Hamilton et al. (2016) propose two quantitative laws of semantic change. First, there is the law of conformity which describes a negative correlation between frequency and semantic change. Second, the law of innovation identifies a positive correlation between polysemy and semantic change. (They evaluate word embeddings against historic changes)

Additionally, Dubossarsky et al. (2015) introduce the law of prototypicality. They establish a negative correlation between prototypicality and semantic change. Prototypicality refers to how representative a word is for its category.

In a follow-up work by Dubossarsky et al. (2017), the validity of the proposed semantic change laws is examined critically. While the results of the previous studies are replicable, the authors conclude that the proposed laws are largely a result of spurious correlation and thus not valid in their strictest form. The effect of word frequency is small, the effect of polysemy can be attributed to collinearity as well as the effect of prototypicality.

Eger and Mehler (2017) apply two graph-based models for corpora across three languages. These include a model consisting of a time series of graphs and a second model where each word is represented by its graph. The authors derive two main findings: (1) word embeddings can be derived from neighbouring words' past embeddings, (2) self-similarity decays linearly over time, therefore the meaning of a word strays away from its original meaning.

Contextual embeddings are word representations that take the context of the words into account. One approach using contextualized word embeddings is the model by Giulianelli et al. (2020). They are among the first to use an unsupervised approach with contextualized word embeddings to identify semantic changes. Specifically, their model uses BERT (Devlin et al., 2018) to obtain word usages, the individual usages are formed into clusters and semantic change is measured with three proposed metrics (entropy difference, Jensen-Shannon divergence and average pairwise distance).

Dynamic embeddings are word representations that consider non-linguistic context, i.e. temporal and spatial features.

Gong et al. (2020) propose an unsupervised model to combine both temporal and spatial attributes in their embeddings. Temporal models aim to identify changes in meaning using time-specific corpora, while spatial information refers to the meaning changes of a word according to the geographic region where it is used. Also, they provide the first dataset to evaluate embeddings across locations.

An approach that combines dynamic and contextual features is presented by Hofmann et al. (2020). Using a pre-trained language model (BERT), they create embeddings that are a function of linguistic (contextual) and extra-linguistic (dynamic) contexts. Additionally, the model maps temporal and spatial features in a joint space, which differs from previous dynamic models.

A common limitation among the discussed papers is that they examine semantic change within a single language. The same holds for works that look at multiple languages as well: the shift for a specific word is not compared across multiple languages. To tackle this, we propose a joint vector space for words across different languages.

Word2Vec (Mikolov et al., 2013) is a method to produce word embeddings. In the skip-gram model, embeddings are computed by predicting neighbouring words using a neural network architecture via backpropagation and stochastic gradient descent. The final embeddings are derived from the weights of the hidden layers.

3 Data

As the basis of our experiments, we used the *HistWords* dataset published by Hamilton et al. (2016). This dataset contains precalculated Word2Vec embeddings of 100,000 words in six different languages, taken from *Google N-grams*, as well as

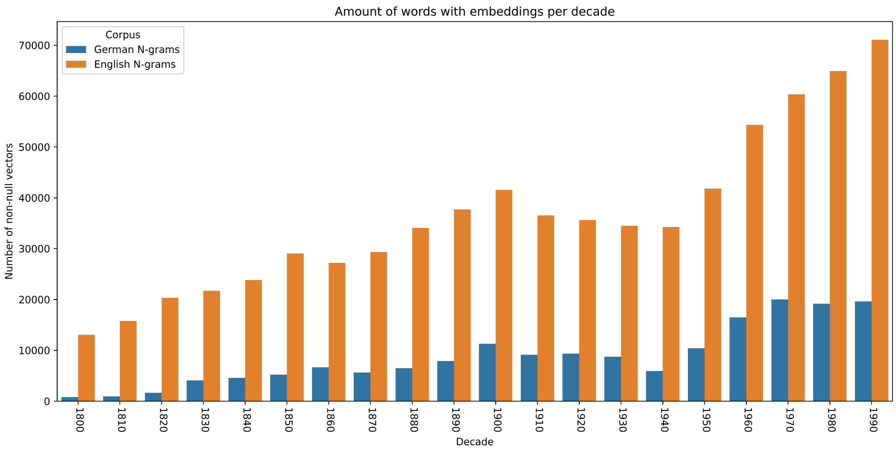

Figure 1: Amount of words in each decade of the different HistWords datasets that have embeddings. Of the $100,000$ words in the dataset, only those that appear more than 500 times in the underlying corpora are included. The other words' embeddings are given as null vectors in the data.

50,000 embeddings from the Corpus of Historical American English (COHA) dataset.

The second edition of the Google Books Ngram dataset, which was used to calculate the embeddings we worked with, was published by Lin et al. (2012). In total, it contains data in eight languages, stemming from 6% of all books ever published (Lin et al., 2012). Since this data was collected using Optical Character Recognition (OCR), it may contain artefacts, i.e. words that were not recognized correctly by the OCR algorithm and are therefore misspelt.

Hamilton et al. use *Word2Vec* to calculate 300-dimensional word vectors based on the 100,000 most frequent words in the Google N-grams dataset. The authors calculate the embeddings for each decade between the 1800s and 1990s for the Google N-grams texts. Discarding words that occur fewer than 500 times within a decade's texts further mitigates possible biases in the sentences, but also reduces the amount of vectors available, in some cases quite substantially. Figure 1 shows the number of non-zero word vectors for each decade.

For our research, we only used the German and English word vectors. To investigate the meaning shifts over time, we divide the data into "old" (1800s to 1890s) and "new" (1900s to 2000s). In order to calculate a single vector representation for each word in the "old" and "new" sets, we used `numpy.nanmean` (Harris et al., 2020) after replacing all null vectors with `numpy.nan`. This ensured that the embeddings' values would not be affected by the years where they did not appear. After calculating the mean, any vectors whose values were still `numpy.nan` were replaced with their original zero values. Table 1 shows how many word embeddings were present in the averaged datasets.

Dataset	Embeddings Old	Embeddings New
English	39308	74758
German	8880	25168

Table 1: Amount of available embeddings after averaging the embeddings for the decades into two time periods. ("Old" meaning 1800s to 1890s, "New" meaning 1900s to 1990s.)

4 Methods

Summary As this is a bilingual study of word meaning change, we first obtain time-specific German and English word embeddings. We then use the VecMap algorithm of Artetxe et al. (2017) to map the four sets of embeddings ("old" and "new" English, "old" and "new" German) into a unique joint space. Through the word embeddings in the joint space, we can analyze the evolution of the relationship between words in different languages over time. Before finding out which word pairs have undergone parallel changes, have drifted apart or converged towards each other across languages over time, we start by choosing the pairs that we want to study. After selecting potential candidates for the convergent, divergent or parallel relative semantic shift (see Figure 2 for a scheme of the classification process), we finally review them manually.

Definitions Two words underwent a *convergent* relative semantic change if they used to have different meanings but acquired a common meaning over the course of time. To illustrate this, we can consider the English word "gay", which used to mean happy or carefree and the German word "schwul", whose meaning possibly was hot or warm. Both words are now a synonym for homosexual. If two words used to have a shared meaning but drifted apart over the years, so that they lost this common meaning or that the link between them has weakened, the relative semantic change is said to be *divergent*. One example of this type of change is the pair "hound" and "Hund" as mentioned previously. Finally, if the word pair used to have a shared meaning, and both words of the pair semantically changed over time and acquired a new shared meaning, the relative semantic change is *parallel*. For instance, "dramatic" and "dramatisch" were initially both related to the field of theatre and drama, but can now also be used to talk about exciting and suspenseful events (definition given by the Collins and the Duden dictionaries respectively).

Embeddings Mapping To create our embedding space we use VecMap, whose principle is to map two sets of word embeddings into a joint vector space. The words can either belong to the same language or different languages. First, we map the time-specific word vectors of the same language into a monolingual joint space, then we combine the mapped word embeddings of the same language in a unique set, and finally, we map all word embeddings of the different languages in the final bilingual joint space. The mappings are done in a semi-supervised way, using small seed dictionaries. We modified the dictionaries several times to aim for the best precision for the mapped embeddings. To visualize the result of the mappings and verify that the mapped embeddings of similar words are near in the joint space, we used the Tensorflow (Abadi et al., 2015) projector tool.

Word Pair Generation To find out which words have undergone parallel changes across languages, semantically converged or semantically diverged, we first need to find relevant bilingual word pairs for which we can try to identify such changes. For this task, our practical approach is to find the k nearest English neighbours of each German word in the joint vector space with respect to their cosine similarity. This way, we obtain k bilingual word pairs per German word. The reason we choose to look for the neighbours of the German words and not the opposite is due to the English embeddings being more precise, which means that the pairs should be more accurate. We also select more than one neighbour as all neighbours did not necessarily undergo the same semantic change, thus some of them may be more relevant for our study than others. To identify divergence, we form the word pairs by using the "old" embeddings to compute the cosine similarity as this change describes words that used to be close. For convergence we use the "new" embeddings, as it describes formerly distant words which became close with time. For the word pairs required for the parallel semantic changes identification, we can choose either way, that is, to form word pairs by using their "old" or their "new" embeddings.

Semantic Change Identification We use the cosine distance of the bilingual word embeddings to evaluate the evolution of the distance between the word pairs in different periods in the joint space to judge whether they may belong to semantic convergence or semantic divergence: if the distance between the pair is much greater in the old period than the distance in the new period, the pair may have undergone semantic convergence, otherwise if it is much smaller than the distance of the new period, it may have undergone semantic divergence. For the detection of semantic parallel changes, we use two thresholds (see section 5), to filter out respectively the cases where the semantic change of the word pair is too small and the ones where the semantic changes of the word pair are not parallel enough. After these two kinds of filtering, the remaining word pairs are the best candidates for the identification of parallel changes across languages. In each case (convergence, divergence and parallel change), we finally manually review the best suiting candidates to make sure they effectively underwent one of the studied semantic changes.

5 Implementation details

Monolingual mappings To map the monolingual time-specific embeddings on the same joint space, we use the semi-supervised setting of the VecMap software. In both cases, we provide the software with a small seed dictionary containing a list of word pairs. This gives the algorithm some examples of how to associate the words of the two sets of word embeddings we want to map. Since

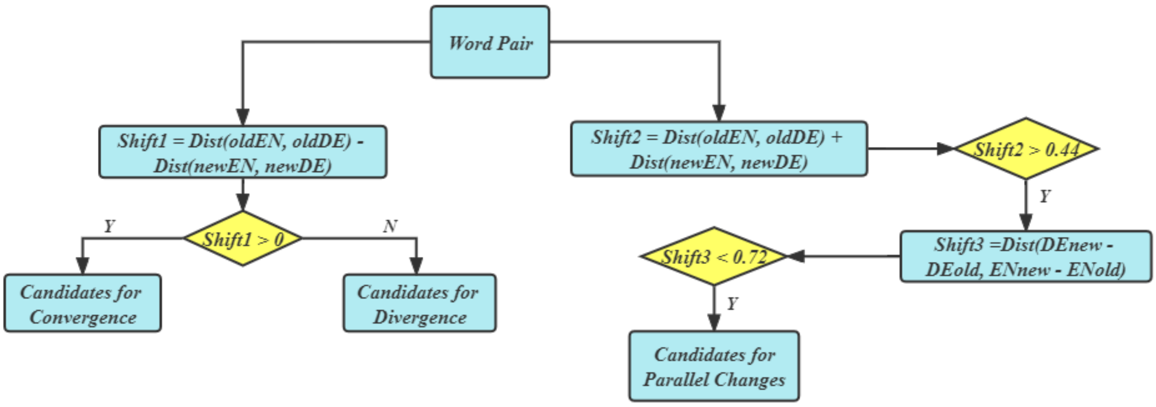

Figure 2: Scheme of the classification process for the word pairs. The candidates identified by the algorithm are sorted before the manual checking to make it more efficient (see section 5).

the goal of the monolingual mappings is to map the "old" and "new" subsets of the same language together, we can use pairs such as apple-apple or cat-cat for the English-English mapping and apfel-apfel or katze-katze for the German-German mapping. For our work, we specifically select words that did not undergo a significant semantic change and pair their "old" and "new" versions. The dictionary for the English-English mapping contains 1163 entries and the dictionary for the German-German mapping contains 513. Both of them comprise pronouns (i-i; du-du), determiners (the-the; der-der), verbs or derived forms of verbs (be-be, doing-doing; sein-sein, wirst-wirst), conjunctions (and-and; oder-oder), prepositions (above-above; unter-unter), adverbs (again-again; einmal-einmal) and the numbers contained in both the "old" and "new" subsets of the languages (512-512; 1872-1872). In the German dictionary, we also sometimes put the same determiners and pronouns in multiple cases: nominative, accusative, dative or genitive (mir-mir, diesen-diesen). Our goal was to be able to generalize with as few entries as possible. We also provide the algorithm with the files containing the "old" embeddings for the language and the "new" embeddings for the language in the Word2Vec format.

Tagging After the two monolingual mappings, we run an algorithm adding a tag to the words in the mapped embeddings files to be able to differentiate them and avoid having duplicate versions of the words in the final bilingual file with all mapped embeddings. For example, apple becomes either apple_oldEN or apple_newEN and apfel becomes either apfel_oldDE or apfel_newDE. We chose to do the tagging only after the first two mappings to prevent it possibly interfering with their outcome and causing the VecMap software to give results of lesser precision. We then merge the files containing the "old" and "new" embeddings to obtain one single file per language that we can use for the final German-English mapping.

Bilingual mapping For the bilingual mapping, we also use the semi-supervised mode of VecMap. The German-English seed dictionary used contains 1249 entries: 379 of them pairing "old" German and English words, and 870 pairing "new" words. For this mapping, the dictionary comprises pronouns (er_oldDE-he_oldEN; wir_newDE-we_newEN), verbs (vergessen_newDE-forget_newEN), adverbs (manchmal_oldDE-sometimes_oldEN), nouns (reptilien_newDE-reptiles_newEN) and the numbers contained in both the "old" or both the "new" subsets of the languages (133_oldDE-133_oldEN; 2004_newDE-2004_newEN). There are some recurrent themes in the dictionaries, but this is only a consequence of the "old" German data being quite limited and these themes having the most representative words in it, for example: time(stunden_oldDE-hours_oldEN; freitag_newDE-friday_newEN), frequency (manchmal_oldDE-sometimes_oldEN; selten_newDE-seldom_newEN), food(butter_oldDE-butter_oldEN; kakao_newDE-cocoa_newEN), raw materials(zink_newDE-zinc_newEN; silber_oldDE-silver_oldEN) or religion (heiland_newDE-redeemer_newEN; evangelium_oldDE-gospel_oldEN). To find which words would fit the best in the dictionary, we use the Tensorflow projector tool to visualize the English and German words in the monolingual embedding spaces. The following criteria were used for the creation of the dictionary: both words of the pair have embeddings that are representative of their semantics well enough, they both did not undergo a significant semantic change (as this is the kind of words we try to find so we cannot force the embedding values to be the same) and they both have a single main meaning (for simplicity purposes). These criteria are assessed by looking at the nearest neighbours with respect to the cosine distance of the "old" and "new" versions of the word and by looking up the words' translation(s) in a bilingual dictionary.

Building bilingual word pairs To find which words are the closest after the bilingual mapping to form the word pairs, we use the cosine similarity of their embeddings. We denote D_e the matrix whose lines are the embeddings of the German words, E_n the matrix whose lines are the embeddings of the English words and we use cos_sim as an abbreviation for the cosine similarity function. We first normalize all embeddings ld_i and le_j of the German and English matrices, and we note $\bar{ld_i}$ and $\bar{le_j}$ their normalized equivalents, to obtain this equation:

$$\begin{aligned} cos\_sim(ld_i,\ le_j) &= \frac{ld_i \cdot le_j}{\|ld_i\| \|le_j\|} \\ &= \bar{ld_i} \cdot \bar{le_j} \end{aligned}$$

This avoids some computation overhead due to calculating the norm of the vectors multiple times. Then, we compute the following matrix multiplication:

$$M = \bar{D_e} \bar{E_n}^T$$

where $\bar{D_e}$ is the matrix of the normalized German embeddings and $\bar{E_n}$ is the matrix of the normalized English embeddings. We obtain:

$$M_{ij} = \bar{ld_i} \cdot \bar{le_j}$$

with $\bar{ld_i}$ being the i^{th} line of $\bar{D_e}$ and $\bar{le_j}$ the j^{th} line of $\bar{E_n}$. As the lines of $\bar{D_e}$ and $\bar{E_n}$ are the normalized German and English embeddings, we finally obtain this result:

$$M_{ij} = cos\_sim(ld_i,\ le_j)$$

where ld_i is the i^{th} line of D_e and le_j the j^{th} line of E_n. We then select the indices of the embeddings of the k English nearest neighbours of a German word with the embedding ld_i:

$$indices_k(ld_i) = argmax_k([M_{ij},\ j \in [0, n]])$$

with n being the number of English words and $argmax_k$ selecting the indices of the k highest values in a list. Finally, the k English neighbours are the words with the embedding le_j such that $j \in indices_k(ld_i)$. For our work, we choose k=10.

This method can be used to either compute the k nearest "old" English neighbours of the "old" version of a German word or the nearest "new" English neighbours of the "new" version of a German word by creating the matrices D_e and E_n with the respective time-specific embeddings.

Identifying parallel changes For identifying parallel changes, we can either start building the word pairs with the set of neighbours computed with the "old" subsets of word embeddings or with the "new" ones. We chose the "new" set of neighbours because the German-English mapping is more precise for the "new" versions of the words, which also means that the neighbours are more accurate. The first indicator assessing if the words of the pair underwent a significant enough semantic change is calculated using the following formula:

$$en\_shift = cos\_sim(oldEN,\ newEN)$$

$$de\_shift = cos\_sim(oldDE,\ newDE)$$

$$mean\_shift = \frac{en\_shift + de\_shift}{2}$$

where oldEN and newEN are the "old" and the "new" embeddings of the English word of the pair and oldDE and newDE are the "old" and the "new" embeddings of the German word.

If the mean_shift value is higher than a first threshold, we then compute the second indicator for the pair, which assesses if the change is parallel enough. The formula used to calculate this indicator is:

$$cos\_sim(newEN - oldEN, newDE - oldDE)$$

The value of this indicator is supposed to be smaller for changes that are more parallel, which is why we compare it to a second maximal threshold.

We iteratively chose both threshold values by first taking the mean value of the semantic change indicator for all pairs of words, then setting the minimal change threshold to this value. After that, we evaluated the mean value of the parallelism indicator and changed the maximal parallelism threshold to this value. We repeated this process until we obtained a desired remaining number of word pairs (around 1000 in our case). The final values we used for the thresholds are:

$$change\_threshold = 0.22$$

$$parallelism\_threshold = 0.72$$

Sorting the word pairs Before we manually review the best-suited word pairs for the convergence, divergence or parallel change, we sort them to make the process easier. For the candidates for convergence and divergence, the sorting criterion is $val = |cos\_sim(oldEN, oldDE) - cos\_sim(newEN, newDE)|$, which characterizes the strength of the change of distance between the two words of the pair over time. We sort by decreasing value, such that the word pairs with the strongest evolution of distance appear first in the list to review. The results given in tables 2 and 3 are also sorted in descending order according to this value. For the candidates for parallel change, we tried both sorting by decreasing value of the semantic change indicator $mean\_shift$ described above and sorting by increasing value of the parallelism indicator. Sorting by the parallelism indicator seems to put more pairs of words that actually underwent a parallel relative semantic change at the beginning of the list.

6 Results

Using the methods described above, we were able to create a list of word pairs for divergence, convergence, and parallel change, respectively. Since all words' semantics are evolving throughout history, there were a lot of results that seem incidental, with the words sharing no obvious commonalities. For example, the two words that diverged the most according to our results were German *wunsch* and English *biederman*, and the words that had the strongest convergence were German *schiff* and English *americana*. While it is possible that these words used to share a similar meaning or do so now, we were unable to discover their connections. This is emblematic of the difficulty we faced, especially concerning divergence: in most cases, we are only aware of the current meaning of words, which makes it hard to identify words that used to be closer in meaning.

Parallel Change Finding a parallel change in the data is difficult for the same reasons, as most of our discoveries seem to be accidental. When we do find a pair of words that can be linked together, their evolution, while parallel, is not very pronounced. Comparing the evolution of the German word *proletariats* and English *proletariat* (see figure 3), for example, indicates that they move perfectly parallel through the embedding space. Even though the change in meaning is subtle, it is supported in the data: in the old texts, proletariat seems to have a more idealized interpretation related to the idea of improving the lives of the working class, with its closest neighbours being *enfranchisement, serf, befreiung* (liberation) and *verbesserung* (improvement). The more modern interpretation seems to be more cynical, with the nearest neighbours being *dictatorship, exploiters, diktatur* (dictatorship) and *klassenkampf* (class struggle). One might expect that the shifts of two words as obviously closely related as proletariat in the two languages will always be parallel, but this is not the case (see figure 6 in the appendix for comparison). This change, therefore, shows that in some cases, the texts discussing a specific topic in two languages evolve in very similar ways throughout time, but more often they do not.

Convergence Some interesting convergence results are displayed in Table 2. One example we want to highlight is the convergent shift in the meanings of English *ministries* and German *ministeriums*, visualized in figure 4. While the German word denotes a governmental office in both the 1800s and the 1900s, illustrated by its nearest neighbours being *innern* (interior as in "ministry of the interior"), *minister* (secretary, minister) and *justiz* (justice) in the old data and *minister*, *staatsministerium* ("state ministry") and *reichsministerium* ("Reich ministry") in the new data. The English *ministries*, however, has undergone a shift: its old meaning used to be "office of one set apart to preach; ecclesiastical function" as given in Dictionary of the English language (Johnson and Walker, 1828). In our data from the 19th century, *ministries* is closest to *fates, subserviency* and *confessedly*. While this meaning is still in use today, our results show that a second interpretation of the word has become more prevalent, which is closer to the meaning of

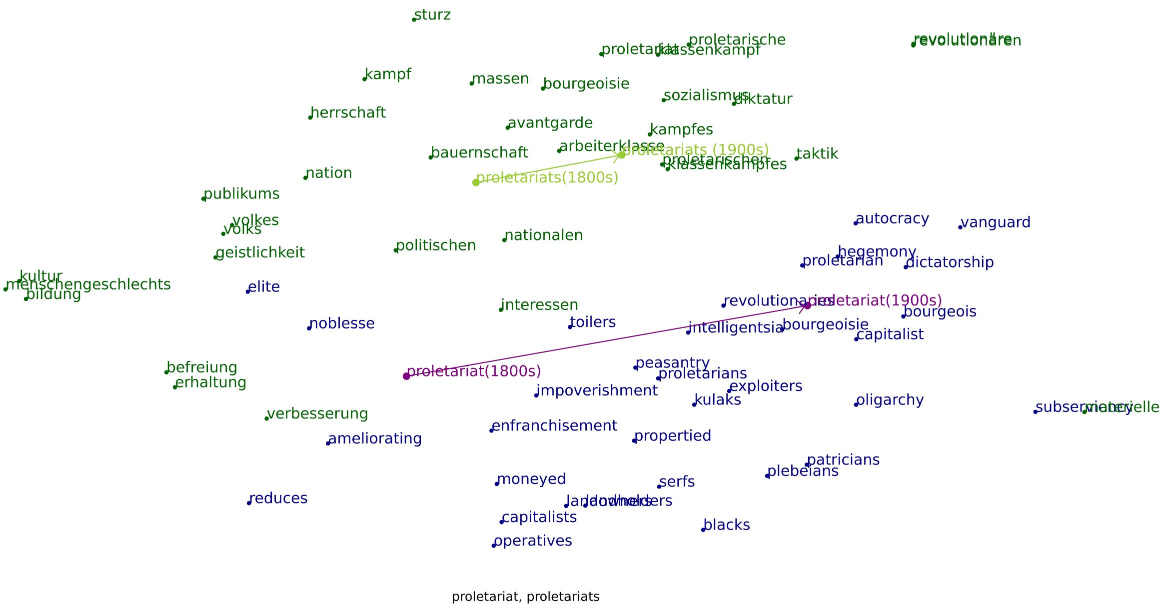

Figure 3: Parallel Change of English *proletariat* and German *proletariats*. This shows that texts dealing with this topic in both languages evolved in similar ways.

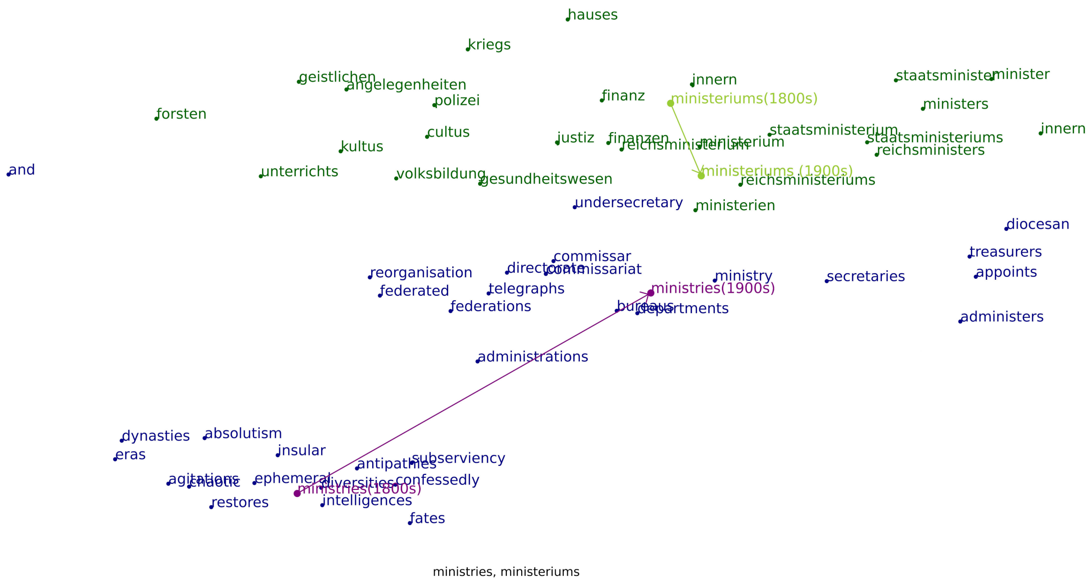

Figure 4: Convergence of English *ministries* and German *ministeriums*. While the German term has always meant a governmental department, the English term used to indicate a clerical office. In the newer data, it has shifted to the government sense as well.

the German word. In the new meaning's vicinity, we find words like *bureaus, commissariat* and *directorate*. This is confirmed by Merriam-Webster describing *ministry* as "the office, duties, or functions of a minister" before "the body of ministers of religion".

Divergence Notwithstanding the difficulties mentioned above, we were able to identify some interesting results for the divergence. One German word that appeared multiple times in the divergence file in connection with English terms connected to mathematics such as *hypotenuse, hyperbola, isosceles* was *axen*. This is likely because *Axen* used to be a customary spelling of what is now spelt "Achsen", meaning axes in English. Looking at the nearest neighbours of *axen* in the old data confirms this: we find terms such as *ebene* (plane), *parallel*, or *senkrecht* (perpendicular). In the data from the 20th century, however, axen denotes GDR politician Hermann Axen, validated by its nearest neighbours

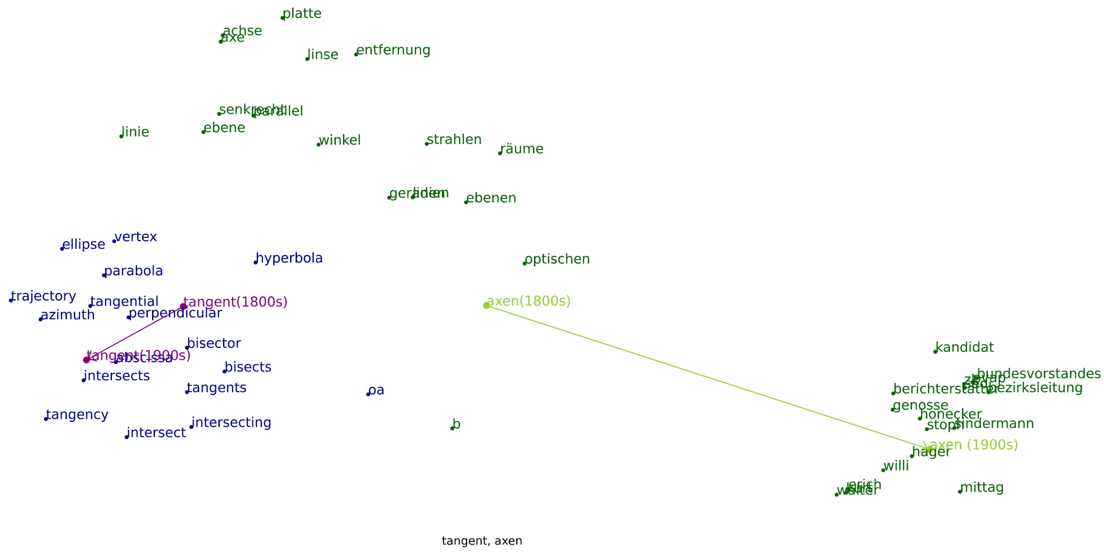

Figure 5: Divergence of English *tangent* and German *axen*. In the 1800s, *axen* used to mean axes, which is spelled *Achsen* in modern days. In the 1900s, it denotes GDR politician Hermann Axen.
We show the evolution of the words' meanings according to the newest embeddings. In case no modern embedding was available for a word, we included its old position in the graph for emphasis.

EN	DE	EN_old	DE_old	Shared meaning (new)
insurgent	*serbischen*	politics	geography	politics
blithedale	*roman*	location	literature	literature
ministries	*ministerium*	clerical	governmental	governmental
factor	*faktor*	trade	mathematics	mathematics
fat	*zuwachs*	biology	economy	biology

Table 2: Word pairs with detected converging semantic shifts. The old context of the words is given in the columns DE_old and EN_old for German and English, respectively. The new, shared meaning is given in the final column.

hager (Kurt Hager, likewise a GDR politician), *genosse* (comrade), and *honecker* (Erich Honecker, leader of the GDR from 1971 to 1989). This shift can be examined in figure 5.

The English *hobbie* and German *froh* (lucky, cheerful) undergo a similar phenomenon, with *hobbie* in the old texts being the equivalent to what is now spelled *hobby*, and therefore being quite close to *froh*. In the 20th century, however, *hobbie* stands for Holly Hobbie, a US-American writer of children's books.

The divergence of English *refugee* and German *priester* (priest) is due to the fluctuating reasons for seeking refuge: nowadays, refugees are identified by their origin, while in the 19th century many people had to flee because of their religion. This explains why *refugee* used to be closer to *priester*.

We have summarized the divergence results in Table 3.

7 Discussion

In this work, we examined and compared the semantic shifts in both English and German, looking for words with diverging or converging meanings, or meanings that evolved in similar, parallel ways.

Limitations The biggest limiting factor in our work was the data. We were unable to find textual German and English datasets that are large enough to produce meaningful embeddings, show little bias concerning the topics, state the publication date of the texts, and cover similar time frames, which would be necessary to make a significant comparison possible.

The *HistWords* dataset we settled on had promised to fulfil all these requirements, offering up to $100,000$ embeddings in both languages over the same time period and ordered by decade. However, due to the threshold stipulation set by the authors, only a small fraction of these $100,000$

EN	DE	Shared context (old)	EN_new	DE_new
hobbie	*froh*	happiness	person (writer)	happiness
refugee	*priester*	religion	origin	religion
transposition	*function*	humanities	biology	mathematics
quadrant	*axen*	geometry	geometry	person (politician)
radius	*axen*	geometry	geometry	person (politician)

Table 3: Word pairs with detected diverging semantic shifts. The shared, old meaning of the word pair is given in the "Shared meaning (old)" column. The new interpretations for English and German are in the EN_new and DE_new columns, respectively.

words actually had embeddings in the published dataset. This greatly restricts the number of worthwhile word pairs we were able to detect. Many terms that would presumably have demonstrated an interesting development, such as *computer* (from a human doing calculations to a machine) did not have embeddings in either the old, new or both datasets. We also had been interested in comparing the earliest files (1800s) to the newest files (1990s), hoping that the shifts would be more distinct than in the averaged files. This was made impossible by the scarcity of German data, especially in the earliest decades: the German 1800s file, for example, contained only 807 embeddings. We also tried to use the *HistWords* embeddings calculated on the basis of the Corpus of Historical American English (COHA, Davies (2012)). Because the COHA texts are hand-selected to be representative of American English, the dataset contains much fewer samples than the Google N-grams, resulting therefore in fewer available embeddings. Of the 50,000 words in the HistWords COHA datasets, only an average of 15,000 embeddings were actually published per decade. In combination with the very sparse German embeddings, this did not allow us to find any meaningful results.

8 Conclusion

In this work, we explore semantic change using a novel approach that examines change for pairs of words across English and German directly. Through mapping embeddings in a joint space and using a cosine-similarity based approach, we are able to show that meaning shifts occur over different languages. In particular, we identify that these shifts exist in both parallel, converging and diverging forms.

Future Work To improve upon our results, better data would be greatly beneficial. A dataset that fulfils the requirements mentioned above would possibly have to be created for this. Access to a large German dataset could make for very interesting and more varied results than ours.

Alternatively, future work could compare English data to a different language with an appropriate dataset that already exists, such as Spanish or French. Being Romance languages instead of a Germanic language like English might produce different, thought-provoking conclusions, like possibly showing more convergence than divergence as the respective countries became closer.

Another interesting possibility could be to compare not the averaged meanings of words for long periods of time, but the data for different decades instead. For example, contrasting the earliest to the latest data could show a stronger shift, whereas studying closer decades might pinpoint when exactly the change happened.

Likewise, using a dataset covering a longer period of time could bring more drastic changes to light. Comparing not only two but more different points in time could even reveal different combinations of the shift patterns we discovered, such as two words that converged at first but then diverged again.

References

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

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.

Leonard Bloomfield. 1933. Language george allen & unwin ltd.

Collins. 2022. dramatic. In *Collins online dictionary*. Collins.

Mark Davies. 2012. Expanding horizons in historical linguistics with the 400-million word corpus of historical american english. *Corpora*, 7(2):121–157.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Haim Dubossarsky, Yulia Tsvetkov, Chris Dyer, and Eitan Grossman. 2015. A bottom up approach to category mapping and meaning change. In *NetWordS*, pages 66–70.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 1136–1145.

Duden. 2022. dramatisch. In *Duden online dictionary*. Bibliographisches Institut.

Steffen Eger and Alexander Mehler. 2017. On the linearity of semantic change: Investigating meaning variation via dynamic graph models. *arXiv preprint arXiv:1704.02497*.

Benjamin W Fortson IV. 2017. An approach to semantic change. *The handbook of historical linguistics*, pages 648–666.

Mario Giulianelli, Marco Del Tredici, and Raquel Fernández. 2020. Analysing lexical semantic change with contextualised word representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3960–3973, Online. Association for Computational Linguistics.

Hongyu Gong, Suma Bhat, and Pramod Viswanath. 2020. Enriching word embeddings with temporal and spatial information. *CoRR*, abs/2010.00761.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. 2020. Array programming with NumPy. *Nature*, 585(7825):357–362.

Valentin Hofmann, Janet B Pierrehumbert, and Hinrich Schütze. 2020. Dynamic contextualized word embeddings. *arXiv preprint arXiv:2010.12684*.

Samuel Johnson and John Walker. 1828. *Johnson and Walker's Dictionary of the English Language*. William Pickering, Chancery Lance, London.

Ekkehard Konig and Johan Van der Auwera. 2013. *The germanic languages*. Routledge.

Yuri Lin, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman, and Slav Petrov. 2012. Syntactic annotations for the google books ngram corpus. In *Proceedings of the ACL 2012 system demonstrations*, pages 169–174.

Merriam-Webster. 2022. Ministry. In *Merriam-Webster.com dictionary*. Oxford University Press.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Appendix

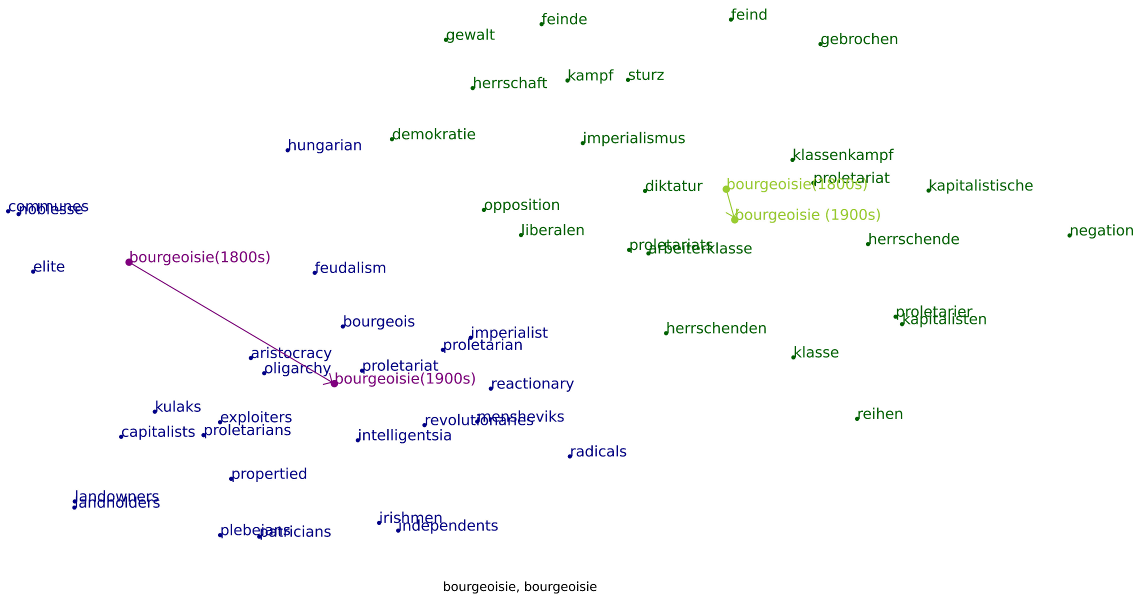

Figure 6: Evolution of the term *bourgeoisie* in both English and German. Even though the words mean the same in both languages, their evolution is not parallel.