

# Tracking Semantic Shifts in German Court Decisions with Diachronic Word Embeddings

Daniel Braun

University of Twente

Department of High-tech Business and Entrepreneurship

d.braun@utwente.nl

## Abstract

Language and its usage change over time. While legal language is arguably more stable than everyday language, it is still subject to change. Sometimes it changes gradually and slowly, sometimes almost instantaneously, for example through legislative changes. This paper presents an application of diachronic word embeddings to track changes in the usage of language by German courts triggered by changing legislation, based on a corpus of more than 200,000 documents. The results show the swift and lasting effect that changes in legislation can have on the usage of language by courts and suggest that using time-restricted word embedding models could be beneficial for downstream NLP tasks.

## 1 Introduction

Languages change over time on different levels, from phonetic and spelling changes to lexical changes, semantic changes, and syntactic changes. Semantic shifts, i.e. changes to the meaning of words, have been researched for hundreds of years, and different taxonomies exist for their classification, e.g. by Bloomfield (1933). These changes are often happening slowly over the course of many years, like the word “dog”, which used to refer to a specific breed and now refers to all breeds (Hollmann, 2009), or the word “broadcast” that in the early 20th century meant “casting out seeds” and now refers to transmitting a signal (Hamilton et al., 2016).

Arguably, the stability of language is higher and more important in legal documents than in most other contexts. Many legal terms, but also terms that are used more freely in everyday language, are well-defined in the context of legal proceedings, either by laws or years of legal precedent. However, language and its meaning and interpretation also change in the context of legal practice. Very prominent examples of that can be found in constitutions. In many western democracies, constitutions

are very stable documents whose texts are hardly changed in decades or sometimes even centuries. Yet, as societies and culture change, the interpretation of these documents by politicians and judges changes as well. Since its implementation in 1949, the first sentence of the second paragraph of article 3 of the German constitution reads “Männer und Frauen sind gleichberechtigt” (Men and women have equal rights). However, it was not until 1977, that a wife would not need the approval of her husband anymore to get a job. A practice that, beyond any doubt, would be ruled unconstitutional today based on the very same sentence. The reasons why the use of language in legal practice can change are manifold, just like in everyday language.

A reason that is particular to the legal domain are changes in legislation that can lead to semantic shifts that are unseen in other domains in speed and thoroughness. By introducing a new law or changing an existing law, legislators have the power to almost instantaneously change the meaning of a word in legal practice. An example of such a quick shift is the German word “Rundfunkbeitrag”. Before 2013, a “Rundfunkbeitrag” was a TV or radio report. However, in January 2013, the “Rundfunkbeitragsstaatsvertrag”(broadcast fee state contract) renamed the German public broadcasting license fee from “Rundfunkgebühr” to “Rundfunkbeitrag”, giving the word a new meaning that quickly has become predominant in legal proceedings (see Section 5). In everyday language, however, the new “Rundfunkbeitrag” is still regularly referred to by its old name.

In recent years, diachronic distributional models, and especially diachronic word embeddings, have been successfully used in different contexts to track semantic changes and changes in language use (see Section 2). The basic idea behind this approach is to train separate word embedding models based on documents from different time periods and then analyse how the word vectors for chosen terms

change between the different models (and hence over time) in relation to other terms (Kutuzov et al., 2018).

This paper presents an application of diachronic word embeddings to the legal domain and specifically focuses on the analysis of semantic shifts introduced by legislative changes. We trained multiple word embedding models on different temporal subsets of a corpus of more than 200,000 decisions from German courts provided by Open Legal Data (Ostendorff et al., 2020), containing texts from the 1970s to 2020.

The results show that diachronic word embeddings can capture immediate and permanent changes in the language used by courts after relevant legislation comes into force. These significant semantic shifts indicate that it could be beneficial for prediction and classification tasks that are connected to words that have undergone such a shift, to use word embeddings that are temporally aligned with this shift, even if that reduces the overall available data.

## 2 Related Work

In 2018, Kutuzov et al. presented a comprehensive survey on works on diachronic word embeddings, including the work of Hamilton et al. (2016), Liao and Cheng (2016), Kutuzov et al. (2017), Rosin et al. (2017), and many others. Therefore, we will focus on works published after 2018.

Kutuzov et al. (2018) found that, as is the case often, most of the existing work focuses on the English language. One notable recent exception from that is the work by Walter et al. (2021), in which the authors analysed a corpus of German parliamentary proceedings spanning from 1867 to 2020. They were, for example, able to show an increase in antisemitic rhetoric in the years leading to the seizure of power by the national socialists.

The data sources that have been used to train diachronic word embeddings in recent years are diverse. Tsakalidis et al. (2021), for example, used a corpus of websites from the UK called DUKweb, Brandl and Lassner (2019) used two newspaper corpora in English and German, and many researchers use the Google Books corpus (Boukhaled et al., 2019; Vijayarani and Geetha, 2020; Yüksel et al., 2021).

While multiple works have been published training diachronic word embeddings on political data and debates, including the previously mentioned

Year(s)	# sentences	Avg. sentences/doc
1970-1979	3,014	26.2
1980-1989	22,102	27.4
1990-1999	170,082	26.4
2000-2009	1,991,396	31.6
2010-2019	4,957,720	38.5
2020	158,705	59.1
1970-2020	7,303,019	36.2

Table 1: Number of sentences and average number of sentences per document per decade

work by Walter et al. (2021), but also work from Rozado and Al-Gharbi (2022) and Indukaev (2021), there is very little work focusing on the legal domain, although the idea that training diachronic distributional models on legal language could provide valuable insights has been voiced before (Rice, 2019).

Soni et al. (2021) were the first, and, as far as we are aware, so far the only ones, to specifically train diachronic word embeddings on court decisions. By training on decisions from US federal courts, they were able to identify decisions that are “on the leading edge of semantic change” and show that such decisions are cited more often. While Soni et al. (2021) focused on semantic change that originates from the decisions themselves and happens at a moderate pace, this paper focuses on semantic change that is introduced to the decisions from an external source, the legislation.

## 3 Data Set

This work is based on the corpus provided by Open Legal Data (Ostendorff et al., 2020). It consists of over 200,000 German court rulings, published between April 2022 and April 1970, by different German courts, ranging from the “Bundesverfassungsgericht” (federal constitutional court) to “Amtsgerichte” (district courts). For reasons of reproducibility, we used the latest stable dump from December 2020<sup>1</sup>, which contains data from April 1970 to December 2020. The dump consists of 201,824 documents from 626 courts.

Figure 1 shows the temporal distribution of the documents and the number of courts that contributed documents in each year. Over time, more and more courts started to publish their decisions digitally and courts also increased the number of

<sup>1</sup><https://static.openlegaldata.io/dumps/de/2020-12-10/>

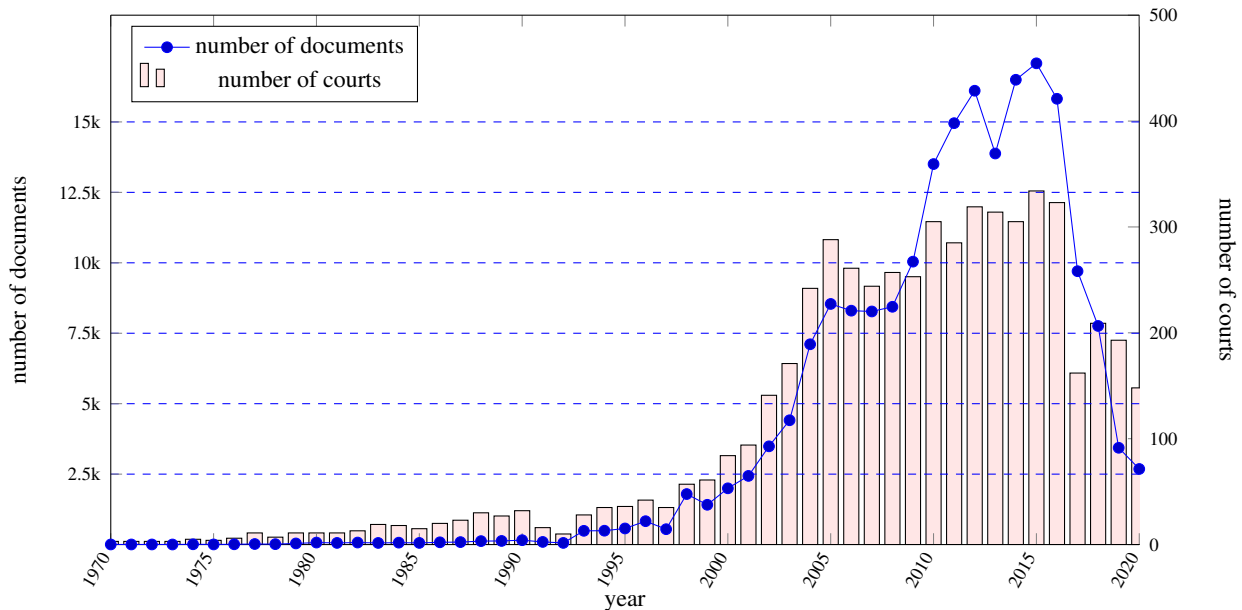


Figure 1: Number of documents per year and the number of courts these documents are from

decisions they published. Therefore, most of the documents in the data set are from recent years. And not only are there more decisions available from more recent years, but the average length of the documents also increased, as is shown in Table 1.

Such a temporal imbalance is very common in legal corpora and the results of this paper suggest, that diachronic word embeddings could help to mitigate the bias that is introduced by such an imbalanced corpus. However, since the available data before 2000 is very limited, with only three to 61 documents per year, we will focus on semantic shifts that happened after 2000.

## 4 Approach

In order to be able to identify the swift semantic shifts we expect to see based on changing legislation, we trained 51 word embedding models, one for each year from 1970 to 2020. That differentiates our approach from works like Hamilton et al. (2016) and Jatowt and Duh (2014), which focus on long-term shifts and therefore aggregate their data over decades. For reasons of comparability and in order to observe more long-term trends, we also trained five models, each of which is spanning a decade from the 1970s to the 2010s. Additionally, we also trained two larger models, one including all data before 2000 and one including all data from 2000 onwards. Lastly, we also trained a model based on all available data, i.e. data from the years

1970 to 2020, resulting in a total of 59 models.

### 4.1 Word Embeddings

We used the Python library `gensim`<sup>2</sup> in version 4.2.0 to train word embeddings with the Word2Vec algorithm (Mikolov et al., 2013). Before starting the training, the data has to be split into individual sentences and tokenized. We used the SoMaJo library (Proisl and Uhrig, 2016) for sentence splitting and tokenization because it has been shown to outperform other libraries on German legal texts (Braun, 2021; Schamel et al., 2022).

We used a window size of five and a vector size of 300 for the word embeddings. The initial learning rate was set to 0.025, the seed to 1, and all words that occurred in the data were included, independent of frequency.<sup>3</sup> Table 2 shows the sizes of the vocabularies for the models spanning a decade, showing a strong correlation between the vocabulary size and the number of sentences that were used to train the model.

### 4.2 Measuring Semantic Shifts

The algorithms that are used to train word embeddings are inherently stochastic, which means they will most likely return different vectors for the same words, even if they are run twice on the exact same data. Therefore, comparing the absolute values of

<sup>2</sup><https://radimrehurek.com/gensim/>

<sup>3</sup>Parameters for the training: `vector_size=300, window=5, alpha=0.025, min_count=1, sample=1e-3, seed=1, epochs=5, workers=4`.

Year(s)	Vocabulary size
1970-1979	20,033
1980-1989	73,310
1990-1999	298,985
2000-2009	1,279,225
2010-2019	2,079,610
2020	204,645
1970-2020	3,322,051

Table 2: Size of the vocabularies of the different word embedding models

word embeddings from different models does not provide meaningful insights. Instead, we want to compare how the position of certain word vectors changes in relation to other word vectors. If, for example, in one model, based on older data, the vector of “to text” is closer to “advertising” than to “smartphone” and in another model, based on newer data, it moves away from “advertising” and closer to “smartphone”, that indicates that the meaning of “to text” is shifting. The similarity of two word vectors can be measured with the cosine similarity. Another approach we use to investigate semantic shifts is looking at the closest neighbours of the word vector of a given word and how these change over time.

Lastly, in order to visualise the change of semantics over the decades in cases that are not related to new legislation, we follow the approach described by [Hamilton et al. \(2016\)](#): For each word that should be analysed, we calculate the union of the word’s  $k$  nearest neighbours in each decade. We then mathematically align and map the different models into a shared two-dimensional system of coordinates, using Principle Component Analysis (PCA). We then plot the vector for the word that we want to analyse in each decade in this system of coordinates. As suggested by [Hamilton et al. \(2016\)](#) we only plot the most “modern” vector for the nearest neighbours, simplifying the plot.

## 5 Results

First, we will look at three examples of rapid semantic shifts that have been caused by changes in legislation (see Section 5.1). We selected three legislative changes which came into force between 2001 and 2013, in order to focus on time periods with sufficient data available. Afterwards, we will look at slower, more traditional patterns of semantic shift in which words change the context they are

used in, based on broader societal changes (see Section 5.2). Finally, we will also take a brief look at changes on the level of the vocabulary (see Section 5.3).

### 5.1 Semantic Shifts caused by Legislation

With the introduction of the “Lebenspartnerschaftsgesetz” (Civil Partnership Act) in 2001 and subsequent changes, the meaning of the word “Lebenspartner” (life partner), as used by courts, shifted gradually from a somewhat vague personal relationship without any legal implications to the meaning of the word “Ehepartner” (spouse) and continued to move closer as the civil partnership received equal rights to the “traditional” marriage in more and more aspects.

Table 3 shows that in the model trained with data from before 2000, the most similar word to “Lebenspartner” is “Lebensgefährten”, which can be seen as a synonym for “Lebenspartner” and also describes a personal relationship without any legal implications. After 2000, the most similar word is “Ehepartner” (spouse), even before “Lebenspartnerin”, the female version of “Lebenspartner”. This clearly indicates a semantic shift that also reflects the new legal implications connected with the word.

Figure 2 shows the cosine similarity between “Lebenspartner” and “Lebensgefährten” and “Ehepartner” for the aligned yearly models. It is notable that right in 2001 when the new legislation was implemented, the cosine similarity between “Lebenspartner” and “Ehepartner” rose significantly. And although there was a drop in the next year, from 2007 onwards, the cosine similarity between “Lebenspartner” and “Ehepartner” was constantly above the similarity of “Lebenspartner” and “Lebensgefährten”.

In 2011, compulsory military service was suspended in Germany, ending 55 years of mandatory service and changing the “Wehrdienst” (military

Rank	1970 - 1999	2000 - 2020
1	lebensgefährten	ehepartner
2	reisepaß	lebenspartnerinnen
3	onkel	lebenspartnerschaft
4	vorgesetzten	ehegatten
5	freunden	lebenspartnern

Table 3: Five most similar words to the term “Lebenspartner” before and after 2000



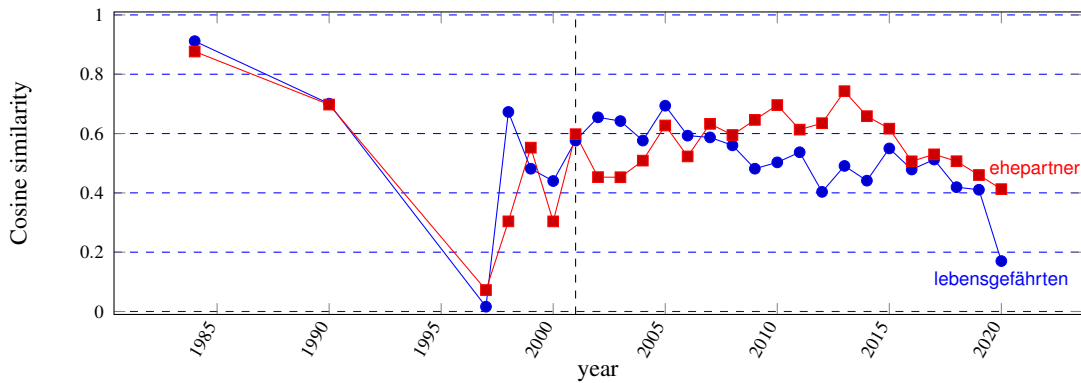


Figure 2: Cosine similarity between the word vector of “Lebenspartner” and “Lebensgefährten” and “Ehepartner” per year

service) from a mandatory to a voluntary service. This change in legislation is also well visible in the corpus of court documents. Before the mandatory service was suspended, the word “Wehrdienst” was closely related to the civilian substitute service (“Zivildienst”, “Ersatzdienst”), as shown in Table 4. While “Militärdienst”, a synonym for “Wehrdienst”, remained the most similar word, the different words describing substitute service disappeared, indicating the shift of “Wehrdienst” from describing a mandatory service to describing a voluntary service.

On a more fine-grained level, an immediate effect is visible in the year 2011 in Figure 4, where the similarity between “Wehrdienst” and “Zivildienst” drops and continues to decline from there, until after 2019 the word is not even in the data anymore.

Rank	1970 - 1999	2000 - 2020
1	militärdienst	militärdienst
2	grundwehrdienst	islam
3	zivildienst	topographen
4	nationaldienst	lienhaushalt
5	ersatzdienst	christentum

Table 4: Five most similar words to the term “Wehrdienst” before and after 2000

In January 2013, the “Rundfunkbeitragsstaatsvertrag” (broadcast fee state contract) renamed the German public broadcasting license fee from “Rundfunkgebühr” to “Rundfunkbeitrag”. Previously, a “Rundfunkbeitrag” would have been a TV or radio report. The new meaning of the word has been adopted very quickly in legal proceedings. Looking at the most similar words in Table 5 reveals an interesting

pattern. Before 2000, all but one entry consist of dates, probably indicating that specific reports are referenced within the documents by the date they have been broadcasted. After 2000, there is a clear connection to the different other fees, as well as the old “Rundfunkgebühr”.

On the yearly level, we can again see how the legislative change has immediate impact on the usage of the “Rundfunkbeitrag” and how it becomes more similar to “Rundfunkgebühr” in 2013.

Rank	1970 - 1999	2000 - 2020
1	07.09	rundfunkgebühr
2	22.12.1980	fremdenverkehrsbeitr.
3	11.8.	kammerbeitrag
4	25.05.1992	kurbeitrag
5	schürfwunde	rundfunkbeitragsstaatsvertrag

Table 5: Five most similar words to the term “Rundfunkbeitrag” before and after 2000

## 5.2 General Context Shifts

In addition to these specific and swift semantic shifts, that are introduced by legislative changes, we can also observe more “classical” semantic shifts in the data, for example, words that are used in different contexts over time. One word for which such a shift has been described often in literature is the word *asylum* (Hamilton et al., 2016; Wiedemann and Fedtke, 2021; Soni et al., 2021). In this corpus, the word “Asyl” (asylum, see Figure 5a) is after 2010 suddenly used very frequently in connection with “Österreich” (Austria) and other European countries, most likely because of refugees seeking asylum reaching Germany through these countries.

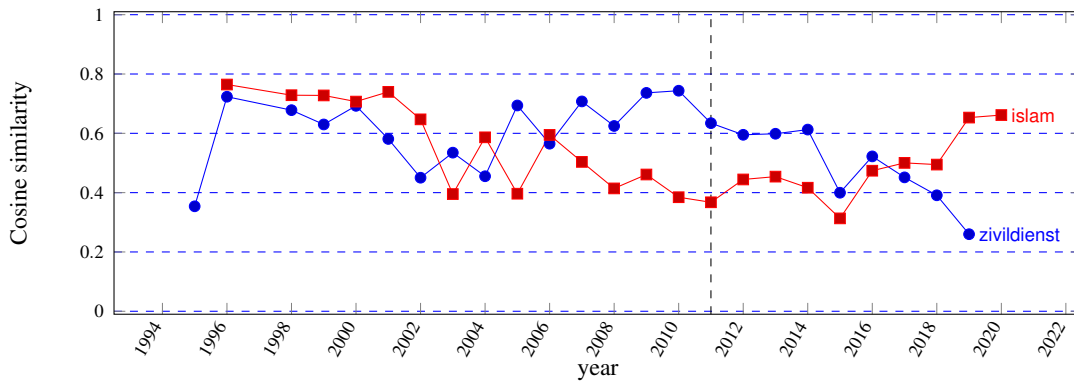


Figure 3: Cosine similarity between the word vector of “Wehrdienst” and “Zivildienst” and “Islam” per year

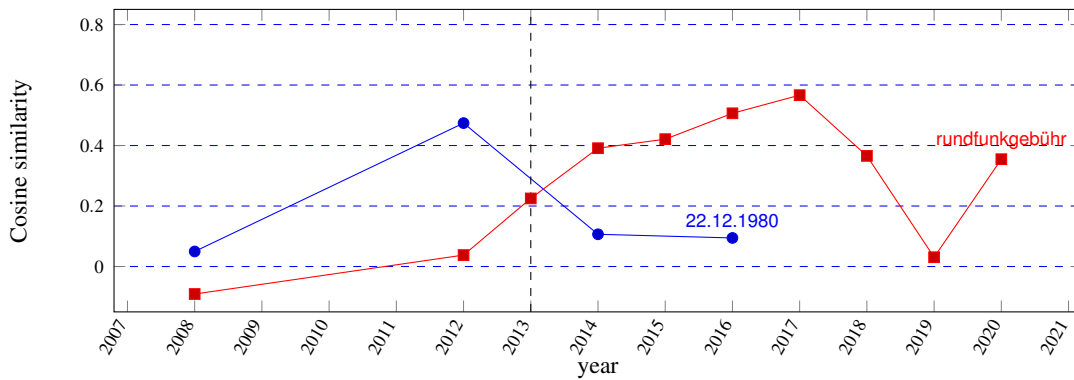


Figure 4: Cosine similarity between the word vector of “Rundfunkbeitrag” and “22.12.1980” and “Rundfunkgebühr” per year

Another shift can be observed for the word “Altlasten”, as shown in Figure 5b. In the 1980s, the word was often used as a euphemism, in the sense of “legacy issues”, to describe the fact that many of the leading figures in German society had already held their positions during the NS rule. Over time, the context in which the word is used, both in the corpus and in society, shifts to contaminated sites or polluted areas, a topic that gains more attention as environmental standards increase.

Other words for which the contexts they appear in have changed include “Freiheit” (freedom, see Figure 5c) and “Geschlecht” (sex and/or gender, see Figure 5d). In our current decade, freedom is in the corpus frequently used in the context of “Versammlungsfreiheit” (freedom of assembly), most likely connected to restrictions of this freedom as part of the measures against the COVID-19 pandemic. For “Geschlecht”, German for both, sex and gender, we can see how it changes from a pure “technicality” (male or female) to a more complex interpretation including health and well-being.

### 5.3 Vocabulary Changes

Since the vocabularies of the 2000s and 2010s are significantly larger than the vocabularies of previous decades, many new words are found in the more recent models, that cannot be found in the older models. However, there is also a number of words and forms of spelling, that are only used in the documents pre-2000s. The German orthography reform in 1996 changed the spelling of numerous words, therefore, variants like “Prozeß” (process or in the legal context also lawsuit) or “wieviel” can only be found in models trained on data from earlier years. In addition, we can also observe words vanishing from the corpus because of changes in legislation, like the disappearance of the word “Zivildienst”, mentioned in Section 5.1.

## 6 Discussion

The results of our analysis show that changes in legislation can cause almost instantaneous semantic changes in the language used by courts and that diachronic word embeddings can be used to track these semantic shifts.

For all instances discussed in the paper, the model that was trained on the complete corpus with data from 1970 to 2020 represented the same meaning as the post-2000 model, which was to be expected, given the temporal imbalance of the data.

If we would want to work on historical decisions, or if a semantic shift would have happened only in 2020, the model trained on all data would most likely misinterpret the words for which a semantic shift has happened. That suggests that using diachronic word embeddings within downstream NLP tasks, like classification or outcome prediction, could be useful in cases where it is known that a word that is important in the context of the task has changed its meaning, e.g. through a change of the law or other significant events. In such cases, aligning the data with such events could help to improve performance, even if it means a reduced corpus for training.

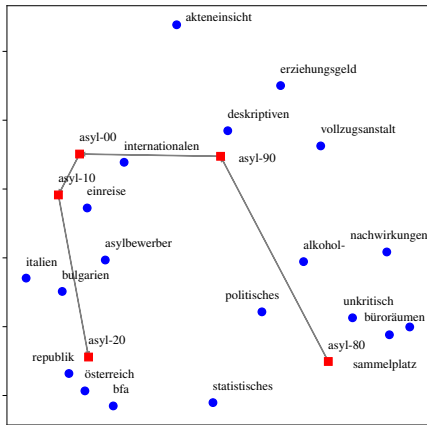
### 6.1 Limitations

The imbalance of the corpus we used for this work limits the generalisability and reliability of the results:

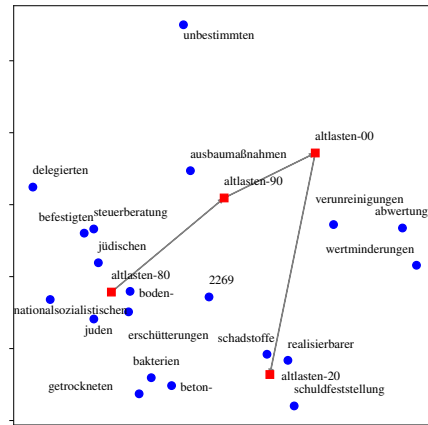
- **Temporal imbalance:** With increasing digitization, the number of available decisions in the corpus also increases, therefore, the data for the 70s and 80s is very limited, affecting both, the comparability, as well as the quality of the word embedding models trained on the data. From 2020 alone, there are already more decisions available in the corpus than from the 70s and 80s combined.
- **Imbalance between courts and court levels:** The highest and higher courts have historically been among the first in Germany to publish their decisions (digitally), therefore, they are over-represented in the dataset, compared to their actual share in decisions made. There is also a difference in the availability of decisions from individual courts and regions, potentially biasing the results. The data from the 70s, for example, contains solely decisions from courts in North Rhine-Westphalia.

## 7 Conclusion

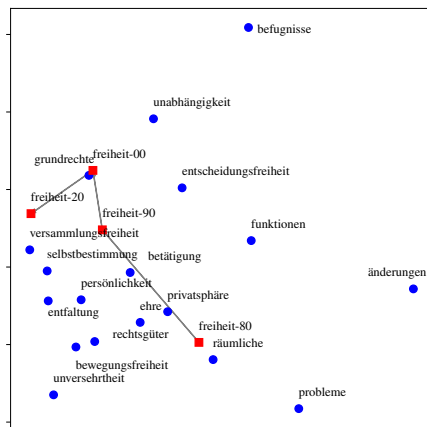
The paper presents an application of diachronic word embeddings to a data set of more than 200,000 German court rulings. 59 different embedding models have been trained, spanning different time spans from years to decades, in order to observe semantic shifts introduced by changes in legislation. The results show that these semantic shifts happen quickly and have a lasting influence on the language that is used by courts.



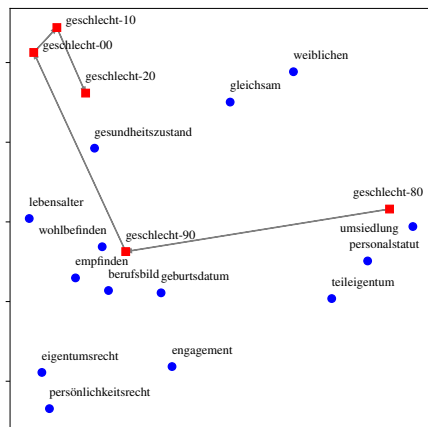
(a) Word "Asyl" (asylum)



(b) Word "Altlasten"



(c) Word "Freiheit" (freedom)



(d) Word "Geschlecht" (sex and/or gender)

Figure 5: Changing contexts in which words are used over time



For tasks, like classification or outcome prediction, that are directly connected to legal terms that have undergone such a semantic shift, it could therefore be beneficial to train word embeddings on a time-restricted dataset, to ensure correct interpretation of the terms in questions, even if that means reducing the available data for training.

In the future, it would be desirable to conduct a similar experiment with a temporally more balanced data set. Another interesting direction for future research would be to connect our findings on the usage of language by courts with the existing literature on semantic shifts in political debates and general language. One could hypothesise that changes in the language used by courts could be predicted by changes in the language used in political debates, which might precede them, and which in turn might be preceded by changes in the general use of language. Analysing whether such influences can be seen in diachronic word embeddings could help to develop models to predict when changed language use will start to have an influence on politics and law.

## References

- Leonard Bloomfield. 1933. *Language*. Allen & Unwin, New York.
- Mohamed Amine Boukhaled, Benjamin Fagard, and Thierry Poibeau. 2019. Modelling the semantic change dynamics using diachronic word embedding. In *11th International Conference on Agents and Artificial Intelligence (NLPinAI Special Session)*.
- Stephanie Brandl and David Lassner. 2019. [Times are changing: Investigating the pace of language change in diachronic word embeddings](#). In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 146–150, Florence, Italy. Association for Computational Linguistics.
- Daniel Braun. 2021. *Automated Semantic Analysis, Legal Assessment, and Summarization of Standard Form Contracts*. Ph.d. thesis, Technische Universität München.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. [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. Association for Computational Linguistics.
- Willem B. Hollmann. 2009. *English Language*, chapter Semantic change. Springer.
- Andrey Indukaev. 2021. Studying ideational change in russian politics with topic models and word embeddings. In *The Palgrave Handbook of Digital Russia Studies*, pages 443–464. Palgrave Macmillan, Cham.
- Adam Jatowt and Kevin Duh. 2014. [A framework for analyzing semantic change of words across time](#). In *IEEE/ACM Joint Conference on Digital Libraries*, pages 229–238.
- 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. Association for Computational Linguistics.
- Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. 2017. [Tracing armed conflicts with diachronic word embedding models](#). In *Proceedings of the Events and Stories in the News Workshop*, pages 31–36, Vancouver, Canada. Association for Computational Linguistics.
- Xuanyi Liao and Guang Cheng. 2016. Analysing the semantic change based on word embedding. In *Natural Language Understanding and Intelligent Applications*, pages 213–223, Cham. Springer International Publishing.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.
- Malte Ostendorff, Till Blume, and Saskia Ostendorff. 2020. [Towards an open platform for legal information](#). In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, JCDL '20*, page 385–388, New York, NY, USA. Association for Computing Machinery.
- Thomas Proisl and Peter Uhrig. 2016. [SoMaJo: State-of-the-art tokenization for German web and social media texts](#). In *Proceedings of the 10th Web as Corpus Workshop*, pages 57–62, Berlin. Association for Computational Linguistics.
- Douglas Rice. 2019. [Understanding legal meaning through word embeddings](#). SSRN 3455747.
- Guy D. Rosin, Eytan Adar, and Kira Radinsky. 2017. [Learning word relatedness over time](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1168–1178, Copenhagen, Denmark. Association for Computational Linguistics.
- David Rozado and Musa Al-Gharbi. 2022. Using word embeddings to probe sentiment associations of politically loaded terms in news and opinion articles from news media outlets. *Journal of Computational Social Science*, 5(1):427–448.

- Tobias Schamel, Daniel Braun, and Florian Matthes. 2022. [Structured extraction of terms and conditions from German and English online shops](#). In *Proceedings of The Fifth Workshop on e-Commerce and NLP (ECNLP 5)*, pages 181–190, Dublin, Ireland. Association for Computational Linguistics.
- Sandeep Soni, Kristina Lerman, and Jacob Eisenstein. 2021. [Follow the leader: Documents on the leading edge of semantic change get more citations](#). *Journal of the Association for Information Science and Technology*, 72(4):478–492.
- Adam Tsakalidis, Pierpaolo Basile, Marya Bazzi, Mihai Cucuringu, and Barbara McGillivray. 2021. [Dukweb, diachronic word representations from the uk web archive corpus](#). *Scientific Data*, 8(1):1–12.
- J Vijayarani and TV Geetha. 2020. [Knowledge-enhanced temporal word embedding for diachronic semantic change estimation](#). *Soft Computing*, 24(17):12901–12918.
- Tobias Walter, Celina Kirschner, Steffen Eger, Goran Glavaš, Anne Lauscher, and Simone Paolo Ponzetto. 2021. [Diachronic analysis of german parliamentary proceedings: Ideological shifts through the lens of political biases](#). In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 51–60.
- Gregor Wiedemann and Cornelia Fedtke. 2021. [From frequency counts to contextualized word embeddings: The saussurean turn in automatic content analysis](#). In *Handbook of Computational Social Science, Volume 2*, pages 366–385. Routledge.
- Arda Yüksel, Berke Uğurlu, and Aykut Koç. 2021. [Semantic change detection with gaussian word embeddings](#). *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3349–3361.