WoVis: Interactive Visualization of Word Embeddings for Semantic Change in Historical and Dialectal Language Resources

Filip Miletić[1] Maximilian Henkel[1,2] Rene Cutura[2] Sophie Sadler[2]
Quynh Quang Ngo[2] Michael Sedlmair[2] Sabine Schulte im Walde[1]
[1]Institute for Natural Language Processing, University of Stuttgart, Germany
[2]VISUS, University of Stuttgart, Germany
[1]{filip.miletic, maximilian.henkel, schulte}@ims.uni-stuttgart.de
[2]{rene.cutura, sophie.sadler, quynh.ngo, michael.sedlmair}@visus.uni-stuttgart.de

Abstract

Computational modeling of language variation and change often relies on comparisons of word embeddings induced from existing historical and dialectal language resources. However, their use in the wider linguistics research community and in application domains such as lexicography is challenged by their limited manipulability for non-technical users, which in turn exacerbates the underuse of such resources. Aiming to foster a broader uptake of embedding-based analyses, we introduce WoVis, an interactive visualization tool designed to compare word embedding models in analyses of semantic change. Our system supports simultaneous model comparisons along two dimensions (e.g., language varieties and time periods) and provides analyses at different levels of granularity: an overview of the full vocabulary across all word embedding models, distributional behavior of individual words, targeted comparisons of word pairs, and model-external lexical features such as frequency and affective norms. We illustrate the utility of our system on two languages, German and English, with analyses of word usage across language varieties as well as time: West vs. East Germany, 1950–1989; and general-domain US vs. scientific UK English, ca. 1800–2000.

Keywords: word embeddings, semantic change, visualization, dimensionality reduction

1. Introduction

Computational modeling of language variation and change often relies on comparisons of word embeddings, i.e., high-dimensional representations of word meaning which reflect cooccurrence patterns in large corpora. Word embeddings are often induced from existing historical and dialectal language resources, and have been used to analyze phenomena including non-standard lexical choices in minority linguistic communities (e.g., Shoemark et al., 2018; Del Tredici et al., 2019; Miletić et al., 2021; Würschinger and McGillivray, 2025); distinctive properties of specialized communication as opposed to the general-domain standard (e.g., Schlechtweg et al., 2019; Bizzoni et al., 2020; McGillivray et al., 2025); and broad trends of semantic change over long periods of time (e.g., Sagi et al., 2009; Gulordava and Baroni, 2011; Hamilton et al., 2016; Schlechtweg et al., 2020; Maurer et al., 2023; Mahdizadeh Sani et al., 2024). But outside of computational research – for example, in linguistics at large and in downstream application domains such as lexicography – the use of embedding-based methods is challenged by their limited manipulability for non-technical users. This issue in turn exacerbates the underuse of existing historical and dialectal language resources, a more general problem particularly affecting lower-resourced languages and language varieties (Soria et al., 2013).

A broader uptake of embedding-based analyses can be supported with visualization systems (Jatowt et al., 2021). This strategy is already incipient in the two-dimensional projections often used to illustrate proposed word embedding analyses, where a target word's trajectory is usually depicted with respect to its nearest neighbors from different models (e.g., Hamilton et al., 2016; Del Tredici and Fernández, 2017; Yao et al., 2018; Rassem et al., 2024; McGillivray et al., 2025). But since these visualizations aim to showcase specific analyses in their language of interest, they are not designed for deployment with other language resources or for interactive exploration – vital factors for novel linguistic insights – nor do they extensively consider the choice of dimensionality reduction techniques. Some of these issues are overcome by general-purpose embedding visualization tools with an interactive interface and multiple dimensionality reductions (e.g., TensorFlow Embedding Projector; Smilkov et al., 2016), but these are not designed for linguistic research and, in particular, lack comparisons of multiple models at the same time. Both analysis-specific and general-purpose visualizations present another important shortcoming – the limited integration of model-external lexical features such as word frequencies and affective properties. These have been shown to correlate with language change (Blank, 1999; Bybee, 2015; Jenkins, 2026) and can therefore aid the interpretation of visualized trends.

Proceedings of Workshop on Dialects in NLP — A Resource Perspective (DialRes) , pages 165–176
16 May 2026.

In order to address these challenges, we provide a series of contributions aiming to improve the accessibility of existing historical and dialectal language resources. **(1) We release WoVis, an interactive visualization tool designed to compare word embedding models** in analyses of semantic change (Figure 1). It supports simultaneous model comparisons along two dimensions (e.g., language varieties *and* time periods) and relies on PaCMAP (Wang et al., 2021), a dimensionality reduction technique which aims to preserve both the local and the global structure of the vector space. The system can be deployed on available word2vec models and also provides a module to train new models from available corpora. **(2) We identify and implement a diverse set of computational semantic analyses** which underpin WoVis, bringing together complementary perspectives on the semantic properties of individual words as well as their relatedness to the rest of the vocabulary. When our system is deployed, it first provides an overview of the full vocabulary across all word embedding models, and then analyzes the evolution of individual words in terms of distributional behavior (nearest neighbors and neighborhood densities), targeted comparisons with other words (capturing shifts in dominant senses), and model-external lexical features (frequency and affective norms). **(3) We show the utility of WoVis with sample analyses of existing resources in two languages**, German and English. We explore a challenging scenario which compares word usage across language varieties as well as time: West vs. East Germany, 1950–1990; and scientific UK vs. general-domain US English, ca. 1800–2000. Importantly, WoVis can also be applied out-of-the-box to compare two word embeddings models from other languages and datasets.[1]

The remainder of this paper first introduces the data (§2) and models (§3) used in our visualizations. We then present the implementation of our system (§4) and the insights from sample analyses (§5). Finally, we close with concluding remarks (§6).

2. Data

This section introduces the data resources underlying our visualizations. While the use of our system is not limited to any specific language or dataset, we showcase its performance on German and English. Both languages are equipped with numerous existing resources, which however may be underused due to limited practical accessibility. We illustrate how our system can help address this challenge, with a particular focus on better utilizing available historical and dialectal resources.

[1]Our code and models are available here: https://github.com/maxeljaxel/wovis

Spiegel		**Berliner Zeitung**	
1950–1959	107.6	1950–1959	400.5
1960–1969	180.8	1960–1969	555.3
1970–1979	228.6	1970–1979	580.2
1980–1989	289.6	1980–1989	618.6
RSC		**CCOHA**	
1800–1849	9.1	1810–1849	44.4
1850–1899	37.0	1850–1899	107.8
1900–1949	65.4	1900–1949	138.6
1950–1996	172.0	1950–2009	181.4

Table 1: Structure of German and English corpora with time periods and size in millions of tokens

For each language, we select two sample corpora which correspond to distinct language varieties and which record their evolution over a parallel time period. Each variety-specific corpus is partitioned into time slices; we refer to these variety–time partitions as *subcorpora*. Their structure is summarized in Table 1 and further described below. Beyond the subcorpora – which are subsequently used to train word embedding models – we further use model-external lexical features to enrich our analysis of target words, also presented below.

2.1. German corpora

Our German-language corpora target the period between 1949 and 1990, during which Germany was formally divided into two states: the Federal Republic of Germany (FRG; West Germany) and the German Democratic Republic (GDR; East Germany). While German was the only official language of both states, substantial political, economic, and social differences led to gradual divergence in language use between the two state-level varieties, especially in the lexicon (Kinne and Strube-Edelmann, 1981; Ammon, 1991; Hellmann, 1993). We aim to capture these divergence phenomena – which for several decades underpinned two competing linguistic standards – by comparing newspaper texts from West and East Germany. For both corpora, we use texts published from 1950 to 1989 and split them into four subcorpora.

Spiegel. As our West German corpus, we use texts from the Spiegel, a weekly news magazine founded in 1947 in Hanover and based in Hamburg since 1952. The magazine mostly covers political and societal topics, with a historical focus on West Germany as well as international events. Our corpus was compiled by crawling the entire digital archive of the Spiegel with permission.[2]

[2]We are grateful to the Spiegel for granting us access to their archive for the purposes of this research: https://www.spiegel.de/spiegel/print/

Berliner Zeitung. As our East German corpus, we use texts from the Berliner Zeitung, a daily newspaper founded in 1945 in East Berlin and published to date. From 1953 to 1990, it was controlled by the Socialist Unity Party of Germany, i.e., the ruling party in East Germany. As a daily newspaper, it has a stronger focus on local events than the Spiegel. We rely on the corpus digitized and published by Berlin State Library.[3]

2.2. English corpora

In our English-language analysis, we compare corpora capturing different national varieties as well as domains (scientific UK English vs. general-domain US English), and we examine them over a longer time span (≈1800–2000). This setup targets the evolution of scientific writing (Biber and Gray, 2016; Teich et al., 2016; Zaberer, 2025) in contrast to the general-domain standard. It is also less restricted than the German-language analysis in terms of time and domains, and as such constitutes a more challenging scenario for our visualization system.

RSC. As a corpus of UK English from the scientific domain, we use the Royal Society Corpus (RSC) (Kermes et al., 2016; Fischer et al., 2020). It contains texts published by the Royal Society of London from 1665 to 1996 in its Philosophical Transactions and Proceedings. We use the texts from 1800 to 1996, split into four subcorpora.

CCOHA. As a corpus of general-domain US English, we use the clean version of the Corpus of Historical American English (CCOHA) (Davies, 2012; Alatrash et al., 2020). It contains texts published between 1810 and 2009, which are moreover balanced across four genres (fiction, non-fiction, magazine articles, and news). We use all available data and once again split it into four subcorpora.

2.3. Model-external lexical features

In order to better explain the semantic change trends we observe, we additionally feed our visualization system with model-external lexical features which characterize individual words and are expected to correlate with semantic change (Blank, 1999; Bybee, 2015). We specifically include word frequency, which we calculate for each subcorpus; and psycholinguistic norms, taken from external resources. For German, we use abstractness, imageability, valence, and arousal norms (Köper and Schulte im Walde, 2016); for English, we use abstractness (Brysbaert et al., 2014) as well as valence, arousal, and dominance norms (Mohammad, 2018). Our visualization system is modular with respect to this type of information, i.e., it can accommodate different or additional lexical features relevant for the visualization task at hand.

[3] https://zefys.staatsbibliothek-berlin.de/ddr-presse/

3. Models

We now turn to the word embedding models used in our visualizations. We discuss the training and alignment procedure; the reduction into two dimensions for visualization; and the computation of neighborhood densities, which we later use to characterize word-level trajectories of semantic change.

3.1. Word embedding models

Following standard practice in computational research on semantic change, we use separate word2vec models (Mikolov et al., 2013) for each subcorpus. We deliberately opt for type-level rather than token-level word embeddings because they are far more compute-efficient and continue to provide robust results in linguistically oriented modeling (recent examples include Mahdizadeh Sani et al., 2024; Miletić and Schulte im Walde, 2025; McGillivray et al., 2025; Würschinger and McGillivray, 2025, i.a.).

For German, we train word2vec models from scratch. We use the skip-gram algorithm with negative sampling, setting window size to 5, vector dimensions to 200, minimum frequency to 2, learning rate to 0.03, and we train for 15 epochs. These hyperparameters were selected after a non-exhaustive grid search evaluated on the German semantic relatedness dataset WordSim280 (Köper et al., 2015).

For English, we rely on previously published models for CCOHA and RSC (Zaberer, 2025), which enables us to implement a resource reuse scenario. The models were trained using the continuous bag-of-words (CBOW) algorithm, with minimum frequency set to 1 and default values for the remaining hyperparameters: windows size of 5, vector dimensions of 100, learning rate of 0.025, and training for 5 epochs.

Once they have been trained, individual word2vec models define embedding spaces whose axes have no absolute meaning and as such cannot be directly compared. To address this issue, we apply the Orthogonal Procrustes method, which aligns two embedding spaces by finding a rotation matrix which minimizes the difference between the spaces while preserving distances and angles (Schönemann, 1966; Hamilton et al., 2016). We individually align all German models to the one trained on the last time slice of the Spiegel corpus, and all English models to the one trained on the last time slice of the CCOHA corpus.

3.2. Dimensionality reduction

Core components of our system visualize the words represented by our word embedding models in a two-dimensional space. We therefore need to reduce our aligned word embedding models from 100 or 200 dimensions (for English and German, respectively) into two dimensions. We do so using Pairwise Controlled Manifold Approximation Projection (PaCMAP; Wang et al., 2021), a dimensionality reduction technique which takes into account neighbor, mid-near, and further pairs of points in order to preserve both the global and the local structure of the vector space. This property is highly relevant for our use case: the global structure enables us to identify broad areas of interest in the lexicon, while the local structure is critical when inspecting the semantic neighborhoods of individual words. PaCMAP is also less sensitive to hyperparameter settings than other widely used techniques (e.g., t-SNE; van der Maaten and Hinton, 2008) and as such is less susceptible to spurious differences in visualization outputs.

3.3. Neighborhood density

Our system provides information on the general semantic properties of individual words by computing neighborhood density, which has been previously used to characterize semantic change over time (e.g., Sagi et al., 2009). We implement two variants of neighborhood density suggested by Schulte im Walde and Frassinelli (2022): **TN**, defined as the average cosine score between a target word t and its k nearest neighbors; and **NN**, defined as the average pairwise cosine score between the k nearest neighbors of a target word t. When observed over time, an increase in neighborhood density (progressively more related nearest neighbors) reflects semantic specialization; and a decrease in neighborhood density (progressively less related nearest neighbors) reflects semantic generalization. Given a user-selected target word t, our system computes the densities on the fly, with the option of interactively adjusting the value of k.

4. WoVis Implementation

We now present the core functionalities of WoVis as designed for deployment on a comparison of two contrastive corpora (e.g., corresponding to two language varieties), each of which is further partitioned into temporal subcorpora. As input, WoVis requires the word2vec models trained on individual subcorpora and subsequently aligned to the same vector space; the dimensionality-reduced representations, which can also be obtained by running the corresponding WoVis module; and an optional file containing supplementary lexical features. See Appendix A for further input requirements.

As shown in Figure 1, WoVis produces visualizations at different levels of granularity: a general overview plot, which shows a two-dimensional projection of all words in all subcorpora (panel A); a neighborhood overview plot, which shows a two-dimensional projection of a selected target word and its nearest neighbors (panel B) together with a table listing the neighbors and their linguistic properties (panel C); overviews of neighborhood evolution for a selected target word, shown in terms of their overlap for all pairs of subcorpora (panel D) and neighborhood densities measured within individual subcorpora (panel E); and a word pair plot, showing the change in relatedness of a target word to user-selected comparison words (panel F). The design of the individual visualizations is explained in detail in the remainder of this section, while the descriptive insights they provide are illustrated using the case studies in Section 5.

General overview The general overview (Figure 1, panel A) uses a scatter plot to show all words from each subcorpus, enabling the user to explore the whole vocabulary and detect clusters of interest. The coordinates of individual points correspond to the two-dimensional PaCMAP reduction of the aligned word2vec embeddings. The plot uses marker shapes and colors to distinguish the corpora (e.g., Spiegel vs. Berliner Zeitung) and brightness to indicate time periods. When hovering over a data point, a label shows the corresponding word and subcorpus. The plot supports zooming into specific areas; brushing, i.e., selecting a subset of points on which to focus; and showing only a selection of subcorpora, with the labels in the legend box serving as toggle buttons.

In a typical second step, the user will move onto an individual word, either by clicking on the corresponding point or by looking it up using the provided auto-complete search box. This action will trigger the neighborhood overview.

Neighborhood overview The neighborhood overview (Figure 1, panel B) appears below the general overview and functions in the same way, with the important exception of the displayed data points. It shows the previously selected target word (using a star-shaped marker) and its k nearest neighbors in each subcorpus, with the value of k set interactively (between 1 and 10,000). This plot enables the user to quickly identify clear differences in meaning across corpora and/or time periods. We also link the two overview plots, showing the cursor location in the neighborhood overview as a crosshair in the general overview; this enables smooth iterations between the two exploration modalities.

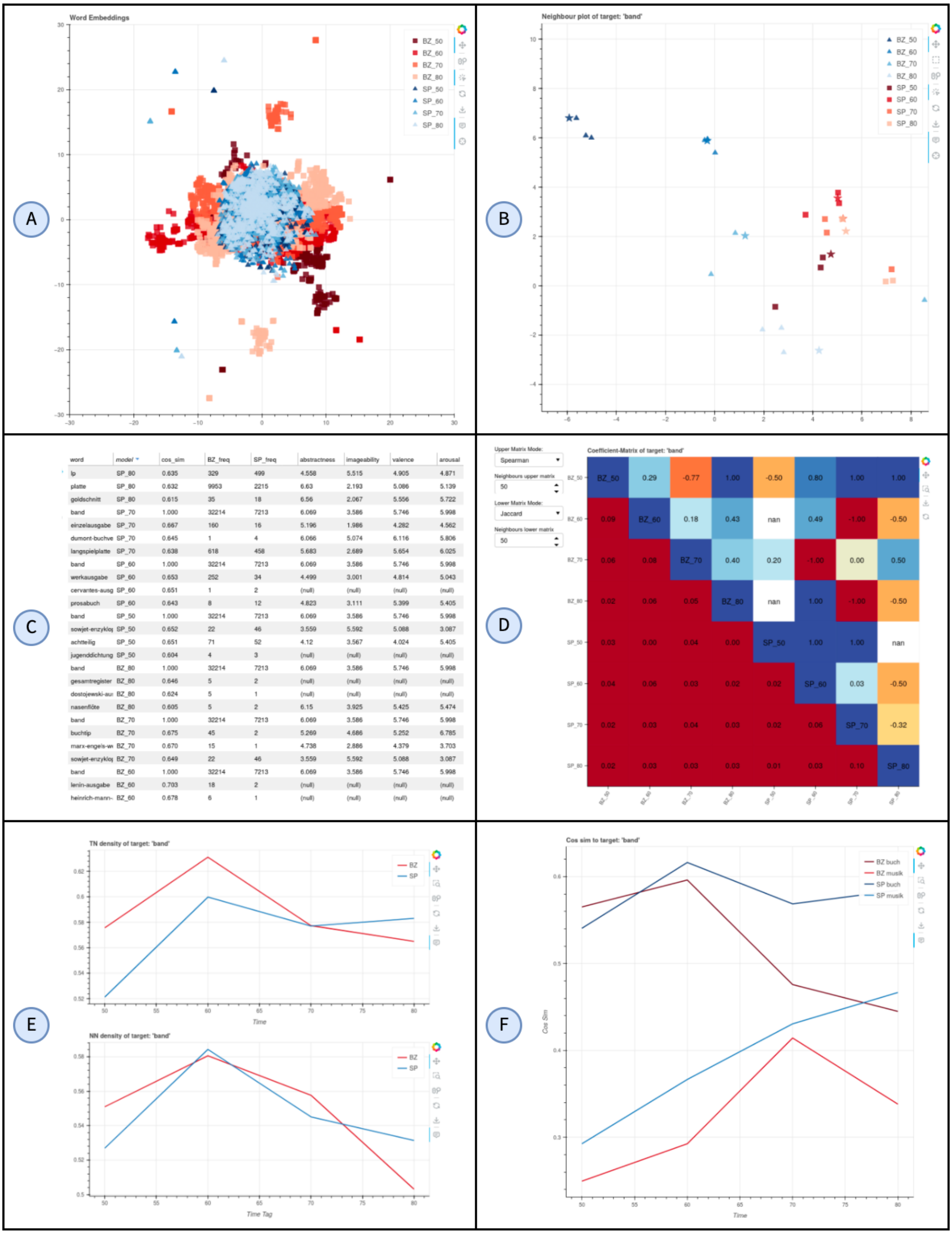

Figure 1: WoVis views for West vs. East German comparison. **A:** general overview; **B:** neighborhood overview; **C:** neighborhood table; **D:** neighborhood overlap over time; **E:** neighborhood density over time; **F:** word pair plot. The shown plots are from our German-language experiments: panels B–E illustrate the neighborhood of the word *Band*; panel F compares it with words *Buch* and *Musik*.

The neighborhood overview also includes a table with the displayed words (Figure 1, panel C), showing the cosine score between the nearest neighbors and the target word as well as any optional model-external lexical features (in our case, frequency and affective norms). A cross-selection between the table and the scatter plot is also supported.

Neighborhood overlap over time As a more structured view of the evolution of a target word's meaning, WoVis provides a heatmap showing the pairwise similarities of the target word's distributional neighborhoods *across* all pairs of subcorpora (Figure 1, panel D). This enables the user to identify pairs of subcorpora with strong shifts in sets of nearest neighbors, which may correspond to points of time in which semantic change occurs.

The neighborhood information is summarized using two measures. The upper triangle of the heatmap shows Spearman's rank-order correlation coefficient, which is calculated based on the cosine scores between the target word and the nearest neighbors; it captures the similarity of nearest neighbors from two corpus sections while accounting for their ordering. The lower triangle of the heatmap shows the Jaccard index, which measures the overlap between two sets of nearest neighbors. This measure is simpler since it ignores the ordering, but potentially more intuitive for the user. The number of nearest neighbors and the position of the two measures in the heatmap can be customized.

Neighborhood density over time To characterize the general semantic properties of a target word *within* a given corpus section, WoVis provides neighborhood density plots (Figure 1, panel E). They show the density value (y-axis) as it evolves over time (x-axis), with line styles used to distinguish the corpora. This view enables the user to identify semantic specialization vs. generalization over time. Plots are provided for both TN and NN density formulations (see Section 3.3), and the number of neighbors can be interactively specified.

The same plot can be toggled to an alternative view showing neighborhood density (y-axis) over the number of neighbors (x-axis), with separate lines representing all subcorpora. This functionality can help determine the useful number of nearest neighbors for a given target word.

Word pair plot As an operational analysis of a target word's dominant senses over time, WoVis provides word pair plots (Figure 1, panel F). These are line plots showing the cosine similarity (y-axis) of the target word to two manually specified reference words, which in a typical scenario reflect two competing senses. The similarity values are plotted over time (x-axis) for each of the two corpora. The user can thus see if the target word has become more or less related to either reference word.

5. Case studies

We now present case studies of semantic change illustrating different insights provided by WoVis.

5.1. Spiegel vs. Berliner Zeitung

We begin with the comparison of German-language corpora: Spiegel, containing texts from West Germany; and Berliner Zeitung, with texts from East Germany. We discuss an example of a lexical process that was highly typical for these two varieties: culture-related English loanwords, which were generally introduced in the West and then spread to East Germany with a delay.

As an example of such a usage, we consider the homonymous word *Band*. In addition to its earliest attested sense 'connection' (neuter; 7th century), it more recently developed the sense 'book volume' (masculine; 18th century) and, borrowing from English, 'musical group' (feminine; 20th century).[4] We mainly focus on the competition between the two more recent senses.

Visualizations for *Band* are shown in Figure 1; we being by observing the neighborhood overview plot (panel B). For Spiegel, the *Band* neighborhood is concentrated in a stable area over the whole time period. For Berliner Zeitung, it is more widespread and moves closer to the Spiegel area over time; this is indicative of progressive convergence. To better understand the senses corresponding to this tendency, we inspect the list of nearest neighbors (Figure 1, panel C; full list in Table 2). In Spiegel, *Band* first appears in music-related contexts in the 1970s, as shown by the nearest neighbor *Langspielplatte* 'long-playing record'. In Berliner Zeitung, music-related nearest neighbors only appear in the 1980s. This indicates a lag in the introduction of the English-borrowed usage in East Germany.

In terms of more general semantic properties, the neighborhood density plot (Figure 1, panel E) shows that density values for both corpora peak in the 1960s and subsequently decrease. The information in this plot subsumes the full range of uses of the target word, so we cannot pinpoint with certainty the most dominant individual sense at a given point in time. However, the overall trend reflects a progressive diversification of contexts in which the target word is attested, which is consistent with the introduction of a new sense after the 1960s.

Bringing together the initial insights, we now set up a word pair plot for the target word *Band* (Figure 1, panel F). We visualize its relatedness

[4] https://www.dwds.de/wb/Band

Spiegel					
1960–1969		1970–1979		1980–1989	
werkausgabe	*collected works*	einzelausgabe	*single issue*	lp	*LP*
cervantes-ausgabe	*Cervantes edition*	dumont-buchverlag	*Dumont Book Publisher*	platte	*record*
prosabuch	*prose book*	langspielplatte	*long-playing record*	goldschnitt	*gilt edge*
silberlöwe	*silver lion*	volkspoesie	*folk poetry*	rockblatt	*rock paper*
subskriptionspreis	*subscription price*	dumont-verlag	*Dumont Publisher*	doppelalbum	*double album*
romanreihe	*novel series*	edition	*edition*	konzert-ausschnitt	*concert excerpt*
einzelausgabe	*single issue*	jubiläumsausgabe	*anniversary edition*	solo-album	*solo album*
werkstattreihe	*workshop series*	originalplatte	*original record*	nachhören	*listen again*
mustervorstellung	*sample presentation*	lp	*LP*	shoes	*shoes*
dtv-band	*dtv volume*	folio	*folio*	gesamtverzeichnis	*full catalog*

Berliner Zeitung					
1960–1969		1970–1979		1980–1989	
lenin-ausgabe	*Lenin edition*	buchtip	*book recommendation*	gesamtregister	*complete index*
heinrich-mann-ausgabe	*Heinrich Mann edition*	marx-engels-werk	*Marx and Engels works*	dostojewski-ausgabe	*Dostoevsky edition*
heine-ausgabe	*Heine edition*	sowjet-enzyklopädie	*Soviet Encyclopedia*	nasenflöte	*nose flute*
werkausgabe	*collected works*	lenin-ausgabe	*Lenin edition*	brennesselsuppe	*nettle soup*
einzelband	*single issue*	urania-universum	*Urania Universum*	beethoven-konzert	*Beethoven concert*
springfrosch	*jumping frog*	wissenschaftsverlag	*scientific publisher*	tanzmusik-sendung	*dance music program*
übersetzung	*translation*	heft	*issue, notebook*	konzertfilm	*concert film*
punktschrift	*Braille script*	essay	*essay*	kunstspiel	*art play*
paperback	*paperback*	goethe-ausgabe	*Goethe edition*	tolstoi-ausgabe	*Tolstoy edition*
edition	*edition*	pierer	*Pierer*	brecht-ausgabe	*Brecht edition*

Table 2: Top 10 nearest neighbors of *Band*, with English translations shown in italics. First decade omitted for space reasons.

(as measured by the cosine score) to two sense-specific reference words: *Buch* ‘book’, indicative of the ‘book volume’ sense; and *Musik* ‘music’, indicative of the ‘musical group’ sense. In both corpora, the relatedness to *Buch* is highest in the 1960s and subsequently decreases. The relatedness to *Musik* also increases: in the Spiegel corpus, from about 0.30 in the 1950s to 0.45 in the 1980s; and in the Berliner Zeitung corpus, from 0.25 to its highest point at 0.40 in the 1970s, then dropping again to 0.35 in the 1980s. These insights pinpoint the trend of *Band* being increasingly used in music-related contexts in both corpora, with a stronger shift in the West than in the East of Germany.

5.2. RSC vs. CCOHA

We now move on to the comparison of English-language corpora: RSC, which represents scientific UK English; and CCOHA, which represents general-domain US English. Since this is a more complex corpus comparison, we discuss two types of variation: differences in word usage across domains and across country-level language varieties.

5.2.1. Variation across domains

As an example of variation in the scientific vs. general domain, consider the case of *gold*. The WoVis-provided nearest neighbors reflect its dominant sense – referring to the precious metal – in both RSC and CCOHA (Table 3). But there are also differences in the contexts in which that sense is used. In RSC, the top nearest neighbors are limited to a broad set of metals (*zinc*, *aluminum*). This points to the technical nature of the underlying contexts, which is further supported by the hyponym *metal* among the nearest neighbors. In CCOHA, the nearest neighbors include other metals (*silver*, *copper*), but also objects made of gold (*jewels*, *coin*) as well as other precious materials used for decorative purposes (*diamond*, *pearl*). This trend indicates more varied, typically everyday, usage contexts in CCOHA.

The qualitative differences observed thus far are also reflected by the quantitative properties computed by WoVis. Neighborhood densities are consistently higher for the RSC than for CCOHA (Figure 2, top), confirming a more restricted set of usage contexts in the scientific domain. We also examine a word pair plot (Figure 2, bottom), selecting as references *iron* (as a type of metal) and *pearl* (as a decorative material). RSC shows clearly higher relatedness for *iron*, and CCOHA for *pearl*; for both corpora, the cosine values are in the 0.7–0.8 range and stable over time. CCOHA also exhibits a moderately strong and stable relatedness with *iron* (≈0.6), indicating that the more technical usage contexts are also attested throughout the time span of the corpus. By contrast, RSC shows limited relatedness with *pearl*, dropping from ≈0.4 to ≈0.2. This finding suggests further specialization of the scientific contexts over time.

RSC			
1800–1849	1850–1899	1900–1949	1950–1996
copper	silver	silver	silver
platina	palladium	zinc	copper
silver	copper	tin	aluminium
tin	platinum	aluminium	nickel
platinum	bismuth	copper	platinum
bismuth	antimony	nickel	tantalum
metal	zinc	platinum	beryllium
zinc	metal	cadmium	tungsten
iron	tin	rhodium	tin
steel	sulphur	graphite	iron

CCOHA			
1810–1849	1850–1899	1900–1949	1950–2009
silver	silver	silver	silver
diamonds	copper	copper	diamond
pearls	pearl	bullion	diamonds
pearl	diamond	pearl	copper
jewels	diamonds	diamonds	rubies
rubies	pearls	diamond	pearl
bracelets	coin	rubies	ruby
diamond	rubies	nuggets	jade
copper	nickel	turquoise	lazuli
coin	jewels	pearls	sapphire

Table 3: Top 10 nearest neighbors of *gold*.

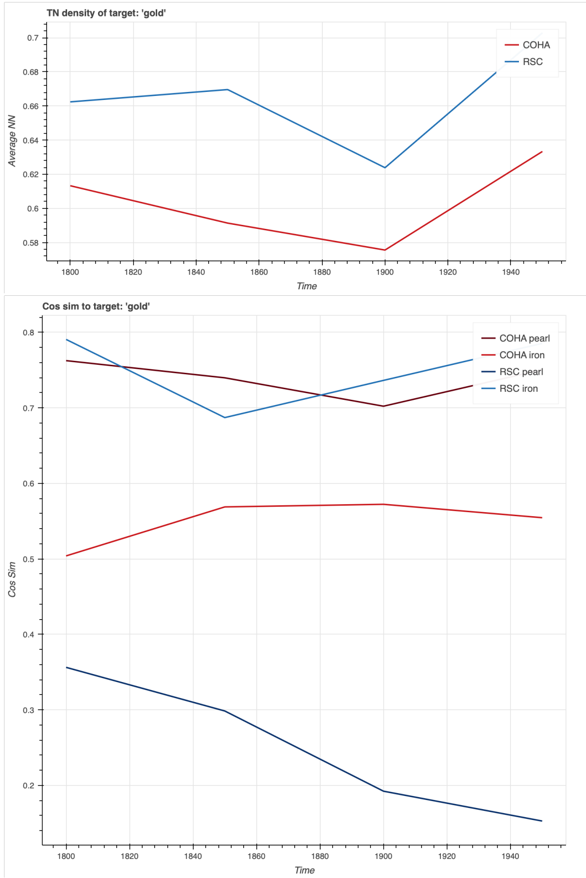

Figure 2: Sample visualizations for *gold* in RSC and CCOHA. Top: TN density (k=100); bottom: word pair plot with references *iron* and *pearl*.

5.2.2. Variation across language varieties

We now look at the ability of our corpora to capture geographic variation in UK vs. US English. We examine the word *gas*, whose dominant sense refers to a state of matter and is attested since the late 18th century.[5] In US (but not UK) English, *gas* is also widely used as an abbreviation of *gasoline*, becoming common by the end of the 19th century.[6]

This characterization finds empirical support in the WoVis lists of nearest neighbors (Table 4). The RSC usage of *gas* corresponds to the earlier sense, as reflected by a diachronically stable set of nearest neighbors: different gases (*hydrogen*, *oxygen*, *chlorine*) as well as other states of matter and related concepts (*liquid*, *fluid*, *air*). CCOHA points to the same nearest neighbors, but importantly also energy sources starting from the 20th century, with prominent nearest neighbors including *gasoline*, *fuel*, and *electricity*.

We once again compare these qualitative differences with the development of neighborhood density (Figure 3, top). CCOHA has a consistently higher neighborhood density, in apparent contrast with the higher degree of polysemy – and hence lower expected density – suggested by individual nearest neighbors. However, it might be the case that the 'gasoline' sense was not only introduced to the US English usage, but also became prevalent to the point of restricting contextual variability. This hypothesis is consistent with the highest density difference between CCOHA and RSC appearing at the beginning of the 20th century, in line with the growing practical significance of gasoline.

We also examine a word pair plot (Figure 3, bottom), selecting as reference words *air*, which references the 'state of matter' sense; and *fuel*, which references the 'gasoline' sense. The plot shows a diachronically stable ordering of dominant senses: the RSC usage is more related to *air* and the CCOHA usage to *fuel*, with the cosine value in both cases ≈0.1–0.3 higher than for the alternative reference word. But beyond this relative difference, both corpora exhibit a diachronic increase in relatedness to *fuel*, consistent with the previously noted differences in neighborhood density.

[5] https://www.etymonline.com/word/gas

[6] https://www.etymonline.com/word/gasoline

RSC			
1800–1849	1850–1899	1900–1949	1950–1996
hydrogen	air	air	liquid
oxygen	gases	gases	gases
chlorine	oxygen	oxygen	air
gases	steam	steam	vapour
vapour	liquid	hydrogen	helium
nitrogen	hydrogen	helium	gaseous
hydrogene	vapour	liquid	reactant
mixture	ozone	nitrogen	solvent
oxygene	argon	vapour	water
phosphorus	chlorine	water	fluid
CCOHA			
1810–1849	1850–1899	1900–1949	1950–2009
acid	carbon	steam	gasoline
fluid	fluid	electricity	fuel
carbonic	boiler	gasoline	electricity
hydrogen	oxygen	helium	steam
oxygen	furnace	oxygen	propane
carbon	vapour	fuel	carbon
liquid	electricity	sulphur	coal
sulphuric	combustion	gases	oxygen
sulphur	wick	alcohol	methane
alcohol	oil	carbon	hydrocarbon

Table 4: Top 10 nearest neighbors of *gas*.

6. Conclusion

We have presented WoVis, an interactive visualization tool for comparisons of word embedding models in semantic change analyses. Explicitly designed to foster novel linguistic insights from existing historical and dialectal language resources, our system supports complex model comparisons capturing two dimensions of variation, e.g., in order to simultaneously contrast language use in different language varieties and time periods. We have deployed WoVis in two scenarios corresponding to the dual-comparison configuration: West vs. East Germany, 1950–1989; and general-domain US vs. scientific UK English, ca. 1800–2000. These case studies demonstrate the utility of our system for research into the evolution of non-standard and domain-specific language varieties, and we hope that WoVis will contribute to a more efficient use of the language resources that target them.

7. Limitations

We illustrated the utility of our visualization method on data for German and English. While the datasets we used correspond to the target category of historical and dialectal language resources, both languages are very highly resourced, which also facilitates access to these kinds of datasets. The overall amount of available language data is considerably smaller for lower-resourced languages and language varieties, which may negatively affect the robustness of the word embedding models trained

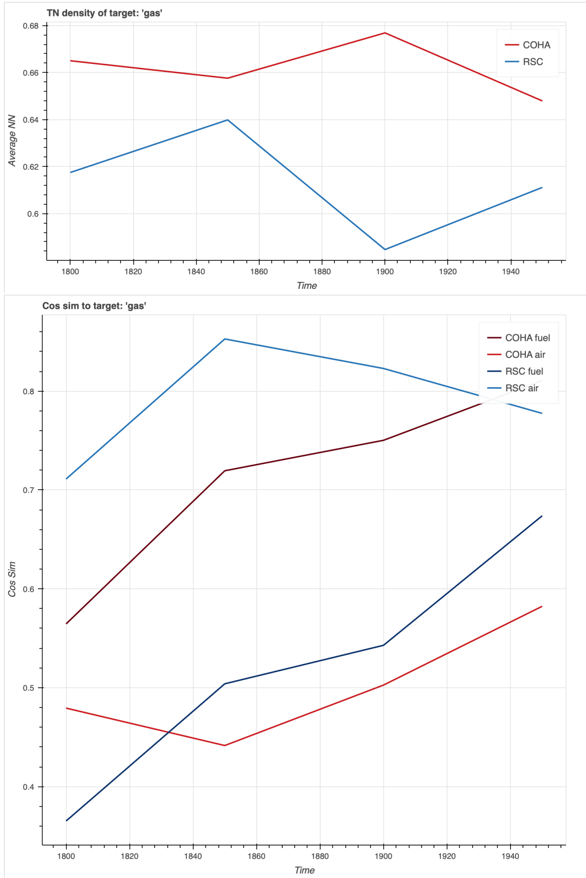

Figure 3: Sample visualizations for *gas* in RSC and CCOHA. Top: TN density (k=100); bottom: word pair plot with references *air* and *fuel*.

on them and consequently limit the informativeness of the visualizations produced by our system.

Moreover, we principally assessed our system through manual inspection of examples of interest. In this way, we simulated system deployment in linguistic research, observing both practical usability as well as potential descriptive insights provided by the visualizations. This approach could be complemented with additional quantitative evaluations (e.g., relying on clustering quality metrics), as well as a more exhaustive set of possible system components (e.g., different word embedding models, dimensionality reduction techniques, and so forth).

8. Acknowledgments

The work presented here was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Research Grant SCHU 2580/5-2 (*Computational Models of Semantic Variation in Multi-Word Expressions across Speakers and Languages*), under Project-ID 251654672 – TRR 161 (project A08), and under Germany's Excellence Strategy – EXC 2120/1 – 390831618.

9. Bibliographical References

Reem Alatrash, Dominik Schlechtweg, Jonas Kuhn, and Sabine Schulte im Walde. 2020. CCOHA: Clean corpus of historical American English. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6958–6966, Marseille, France. European Language Resources Association.

Ulrich Ammon. 1991. The differentiation of the German language into national varieties of the Federal Republic of Germany (F.R.G.), the German Democratic Republic (G.D.R.), Austria and Switzerland. *History of European Ideas*, 13(1-2):75–88.

Douglas Biber and Bethany Gray. 2016. *Grammatical Complexity in Academic English: Linguistic Change in Writing*, 1 edition. Cambridge University Press.

Yuri Bizzoni, Stefania Degaetano-Ortlieb, Peter Fankhauser, and Elke Teich. 2020. Linguistic variation and change in 250 years of English scientific writing: A data-driven approach. *Frontiers in Artificial Intelligence*, 3:73.

Andreas Blank. 1999. Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. In Andreas Blank and Peter Koch, editors, *Historical Semantics and Cognition*, pages 61–90. De Gruyter.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3):904–911.

Joan Bybee. 2015. *Language Change*. Cambridge University Press.

Mark Davies. 2012. Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora*, 7(2):121–157.

Marco Del Tredici and Raquel Fernández. 2017. Semantic variation in online communities of practice. In *IWCS 2017 – 12th International Conference on Computational Semantics – Long Papers*.

Marco Del Tredici, Raquel Fernández, and Gemma Boleda. 2019. Short-term meaning shift: A distributional exploration. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2069–2075, Minneapolis, Minnesota. Association for Computational Linguistics.

Stefan Fischer, Jörg Knappen, Katrin Menzel, and Elke Teich. 2020. The Royal Society Corpus 6.0: Providing 300+ years of scientific writing for humanistic study. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 794–802, Marseille, France. European Language Resources Association.

Kristina Gulordava and Marco Baroni. 2011. A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 67–71, Edinburgh, UK. Association for Computational Linguistics.

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.

Manfred W. Hellmann. 1993. Ostdeutsch - Westdeutsch im kontakt. Brücke oder schranke der verständigung? *Germanistische Mitteilungen*, 38:3–35.

Adam Jatowt, Nina Tahmasebi, and Lars Borin. 2021. Computational approaches to lexical semantic change: Visualization systems and novel applications. In Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen, editors, *Computational Approaches to Lexical Semantic Change*. Language Science Press, Berlin.

Christopher Jenkins. 2026. *Compound(ing) Interest in Interesting Compounds or, Semantic Change and Distributional Representation of Noun Compounds*. Ph.D. thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Hannah Kermes, Stefania Degaetano-Ortlieb, Ashraf Khamis, Jörg Knappen, and Elke Teich. 2016. The Royal Society Corpus: From uncharted data to corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1928–1931, Portorož, Slovenia. European Language Resources Association (ELRA).

Michael Kinne and Birgit Strube-Edelmann. 1981. *Kleines Wörterbuch des DDR-Wortschatzes*, 2. aufl edition. Schwann, Düsseldorf.

Maximilian Köper, Christian Scheible, and Sabine Schulte im Walde. 2015. Multilingual reliability and "semantic" structure of continuous word spaces. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 40–45, London, UK. Association for Computational Linguistics.

Maximilian Köper and Sabine Schulte im Walde. 2016. Automatically generated affective norms of abstractness, arousal, imageability and valence for 350 000 German lemmas. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2595–2598, Portorož, Slovenia. European Language Resources Association (ELRA).

Samin Mahdizadeh Sani, Malak Rassem, Chris Jenkins, Filip Miletić, and Sabine Schulte im Walde. 2024. What can diachronic contexts and topics tell us about the present-day compositionality of English noun compounds? In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17449–17458, Torino, Italia. ELRA and ICCL.

Maximilian Maurer, Chris Jenkins, Filip Miletić, and Sabine Schulte im Walde. 2023. Classifying noun compounds for present-day compositionality: Contributions of diachronic frequency and productivity patterns. In *Proceedings of the 19th Conference on Natural Language Processing (KONVENS 2023)*, pages 40–51, Ingolstadt, Germany. Association for Computational Lingustics.

Barbara McGillivray, Nilo Pedrazzini, Arianna Ciula, Jon Lawrence, Tiffany Ong, Mia Ridge, and Jose Miguel Monteiro Vieira. 2025. Analysing the language of mechanisation in nineteenth-century British newspapers. In Ruth Ahnert, Emma Griffin, and Jon Lawrence, editors, *Living with Machines*. University of London Press.

Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, ICLR 2013, Workshop Track Proceedings*, Scottsdale, Arizona, USA.

Filip Miletić, Anne Przewozny-Desriaux, and Ludovic Tanguy. 2021. Detecting contact-induced semantic shifts: What can embedding-based methods do in practice? In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10852–10865, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Filip Miletić and Sabine Schulte im Walde. 2025. Modeling the evolution of English noun compounds with feature-rich diachronic compositionality prediction. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 20071–20092, Vienna, Austria. Association for Computational Linguistics.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, Melbourne, Australia. Association for Computational Linguistics.

Malak Rassem, Myrto Tsigkouli, Chris W. Jenkins, Filip Miletić, and Sabine Schulte im Walde. 2024. Visualising changes in semantic neighbourhoods of English noun compounds over time. In *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, pages 240–246, Miami, USA. Association for Computational Linguistics.

Eyal Sagi, Stefan Kaufmann, and Brady Clark. 2009. Semantic density analysis: Comparing word meaning across time and phonetic space. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, pages 104–111, Athens, Greece. Association for Computational Linguistics.

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

Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. SemEval-2020 task 1: Unsupervised lexical semantic change detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1–23. International Committee for Computational Linguistics.

Peter H. Schönemann. 1966. A generalized solution of the Orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10.

Sabine Schulte im Walde and Diego Frassinelli. 2022. Distributional measures of semantic abstraction. *Frontiers in Artificial Intelligence*, 4:796756.

Philippa Shoemark, James Kirby, and Sharon Goldwater. 2018. Inducing a lexicon of sociolinguistic variables from code-mixed text. In *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*, pages

1–6, Brussels, Belgium. Association for Computational Linguistics.

Daniel Smilkov, Nikhil Thorat, Charles Nicholson, Emily Reif, Fernanda B. Viégas, and Martin Wattenberg. 2016. Embedding Projector: Interactive visualization and interpretation of embeddings. In *Proceedings of NIPS 2016 Workshop on Interpretable Machine Learning for Complex Systems*. arXiv.

Claudia Soria, Joseph J Mariani, and Carlo Zoli. 2013. Dwarfs sitting on the giants' shoulders - how LTs for regional and minority languages can benefit from piggybacking major languages. In *Conference of the Foundation for Endangered Languages*, Ottawa, Canada.

Elke Teich, Stefania Degaetano-Ortlieb, Peter Fankhauser, Hannah Kermes, and Ekaterina Lapshinova-Koltunski. 2016. The linguistic construal of disciplinarity: A data-mining approach using register features. *Journal of the Association for Information Science and Technology*, 67(7):1668–1678.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605.

Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. 2021. Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization. *Journal of Machine Learning Research*, 22(201):1–73.

Quirin Würschinger and Barbara McGillivray. 2025. Semantic change and socio-semantic variation: The case of COVID-related neologisms on Reddit. *Linguistics Vanguard*, 11(s3):339–352.

Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. 2018. Dynamic word embeddings for evolving semantic discovery. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18, pages 673–681, New York, NY, USA. Association for Computing Machinery.

Urs Zaberer. 2025. Noun compound compositionality in scientific English. Master's thesis, University of Stuttgart, Stuttgart, Germany.

A. Input Format

For each subcorpus being visualized (e.g., Berliner Zeitung 50s, Berliner Zeitung 60s, etc.), WoVis requires the following input files:

- A trained word2vec model in the Gensim KeyedVectors format together with the corresponding NumPy file;
- The corresponding two-dimensional word and coordinate NumPy file, which can also be produced by running the requisite WoVis module;
- An optional CSV file containing model-external lexical features (e.g., word frequency and affective norms), with the first column containing the target words, and the header containing the labels for the included information.

All file names consist of two components: a corpus name tag and a time tag, separated by an underscore (e.g., BZ_50). This information is used when rendering the plots. The name tag determines the general base color in all plots and the marker shape in scatter plots, while the time tag controls the brightness, with more recent data represented using lighter shades. Further implementation details are provided in WoVis documentation.