*Natural Language Processing* (2026), **32**, pp. 233–266
doi:10.1017/nlp.2026.10023

CAMBRIDGE UNIVERSITY PRESS

SURVEY PAPER

# Survey in characterizing semantic change

Jader Martins Camboim de Sá[1,2], Marcos Da Silveira[1] and Cédric Pruski[1]

[1]Luxembourg Institute of Science and Technology, Esch-sur-Alzette, L4362, Luxembourg and [2]University of Luxembourg, Esch-sur-Alzette, L4365, Luxembourg
**Corresponding author:** Jader Martins Camboim de Sá; Email: jader.martins@list.lu

(Received 25 January 2024; revised 28 December 2025; accepted 9 March 2026; first published online 23 April 2026)

**Abstract**

Living languages continuously evolve to reflect the cultural changes of human societies. This evolution manifests through neologisms (new words) or the **semantic change** of existing words (new meanings for existing words). Understanding the meaning of words is vital for interpreting texts from different cultures (regionalisms or slang), domains (e.g., technical terms), or time periods. In computer science, this phenomenon is relevant to computational linguistics tasks such as machine translation, information retrieval, and question answering. Semantic change can impact the performance of these applications, making it important to understand and characterize these changes formally. This problem has recently attracted significant attention from the computational linguistics community. Several approaches can detect semantic changes with good precision, but more effort is needed to characterize *how* word meanings change and to determine how to mitigate the impact of this phenomenon. This survey provides a comprehensive overview of existing approaches to the **characterization of semantic change**. We also formally define three classes of characterization: change in dimension (whether a word's meaning becomes broader or narrower), change in orientation (whether a word acquires a more pejorative or ameliorative sense), and change in relation (whether a word is used in a new figurative context, such as a metaphor or metonymy). We demonstrate the applicability of this formalism on existing corpora, summarize the key aspects of selected publications, and discuss current needs and trends in research on semantic change characterization.

**Keywords:** lexical semantic change; semantic shift; word sense disambiguation

## 1. Introduction

Language is one of the most dynamic skills humans have acquired, constantly evolving to adapt communication to new purposes and needs (Pinker, 2003, p 13–23). While no single unified theory explains its origins (Allan, 2013, p 13–52), language has clearly undergone significant modifications throughout human history. One of the prevailing theories of language evolution is based on the assumption of '*continuity* of change' (Rilling *et al.* 2012).

This theory posits that language, being too complex to have appeared suddenly, must have evolved gradually from earlier pre-linguistic systems used by our ancestors (Pinker and Bloom 1990, p 725–726). In this context, comprehending language change contributes to the fields of history, sociology, linguistics, philosophy, and Natural Language Processing (NLP) (Campbell, 2013, p 1–8), as it is essential to understand the origins of language, cultural development, grammatical theories, and the relationship between language and thought (Sherwood, Subiaul, and Zawidzki 2008, p 426–454).

Language change is driven by many unplanned factors, such as linguistic evolution, sociocultural and historical influences, and technological advancements. A variety of grammatical factors also promote changes, including:

- **Phonological changes:** Changes in the pronunciation and sound systems of a language. For instance, the pronunciation of vowel sounds changed significantly due to the Great Vowel Shift in Medieval English (Nevalainen and Traugott 2016, p 367).
- **Orthographic changes:** Alterations in spelling conventions, often reflecting phonological changes, for example, Latin *panem* to French *pain* ('bread').
- **Syntactic changes:** Changes in grammar and word order. For example, Modern English emphasizes word order and auxiliary verbs more than the intricate inflectional system of Old English (Nevalainen and Traugott 2016, p 616).
- **Semantic changes:** Words or phrases acquiring new meanings over time, for example, *notebook* from 'a book for notes' to 'a portable computer'.

While phonological and morphological changes are crucial, semantics plays a pivotal role in language evolution. Semantic change reflects how words adapt to evolving cultural, technological, and societal contexts, creating new meanings and preserving language's fundamental purpose of effective communication. Despite being a recurring topic in scientific research, semantic change remains less understood than other forms of linguistic evolution due to its often subtle nature compared to more observable shifts, like those resulting from grammar reforms (Allan and Robinson 2012, p 2–3).

Historically, the study of semantic change relied on manual analysis. Linguists carefully examined historical texts, dictionaries, and corpora to track how word meanings evolved. While these studies provided valuable insights, they were slow, labor-intensive, and difficult to replicate.

The advent of computational linguistics and advanced models marked a major shift in the field. With vast digital corpora and efficient algorithms, researchers began to apply statistical methods from NLP to analyze text automatically. This marked a transition from purely observational methods to data-driven statistical analysis (Allan and Robinson 2012, p 161–164).

With access to large text corpora and powerful algorithms, linguists can now detect subtle patterns in language with greater accuracy and efficiency. This allows for the study of languages with limited prior linguistic knowledge (Conneau *et al.* 2018; Wu and Dredze 2019), given sufficient data, and enables comparison of multiple languages to test for universal properties in latent space (Vulic, Ruder, and Søgaard 2020; Wang, Wu, and Neubig 2022).

By adopting a model-based approach, linguists can now explore complex language phenomena across different cultures and backgrounds (Nivre *et al.* 2020). These computational tools provide a structured way to test hypotheses, refine mathematical models, and validate findings against real-world linguistic data (Hammarström *et al.* 2019).

While initially a topic of theoretical interest, the computational study of semantic change has intrinsic value for understanding language evolution and supporting linguistic research. Beyond its linguistic importance, it also informs certain computer science tasks that involve historical language or evolving terminology. For example, historical knowledge graphs such as Wikidata need to account for changes in entity and relation meanings over time to maintain consistency (Polleres *et al.* 2023). In digital humanities, tracking the semantic shift of key concepts (e.g., liberal, gay, computer) over centuries enables a more accurate interpretation of historical texts. Similarly, domain-specific information retrieval can benefit from awareness of semantic drift – for example, the term cloud in technical literature has shifted significantly over the past decades.

Developing cost-effective tools that can handle linguistic variation and precisely characterize semantic changes is crucial for many applications. While the phenomena causing semantic change have been studied extensively in Linguistics (Traugott, 2017), providing deep analyses and typologies, and considerable effort has been devoted to *detecting* such changes by the computational

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

linguistics community (Tahmasebi, Borin, and Jatowt 2018), to the best of our knowledge, few efforts have focused on formalizing the problem to characterize the *type* of change a word has undergone (Hengchen *et al.* 2021).

A formal, mathematical language can serve as a valuable tool in the computational study of semantic change. While we acknowledge that language evolution is deeply intertwined with cultural, societal, and historical dynamics – factors that often resist formalization – developing a structured framework allows us to isolate and investigate specific computational aspects of this complex phenomenon. Our aim is not to reduce semantic change to purely formal terms, but to provide a simplified, tractable representation that supports the development of methods and evaluation tools within computational linguistics. We recognize that such formalizations are necessarily reductive and do not capture the full richness of linguistic and cultural processes; rather, they offer a complementary perspective that can contribute to broader, interdisciplinary understandings of meaning change.

This survey addresses the need to understand and formalize the characterization of semantic change. Characterizing such changes will open new perspectives, particularly since we often lack the resources to study them in detail.[a] This is especially problematic in contexts where multiple word meanings are relevant, potentially leading to political or social misinterpretations. This survey's objectives include:

(1) Identifying the limitations of current representations and methodologies in categorizing semantic changes.
(2) Evaluating the strengths of approaches used to classify Lexical Semantic Change (LSC).
(3) Finding and analyzing formal definitions based on first-order logic to reduce ambiguity and facilitate the comparison of findings.
(4) Demonstrate the task of the LSC conceptually.

In this survey, we select categories that are both addressable by computational methods in a comparative setting (Tang, Qu, and Chen 2013) and well-established in the linguistic literature (Campbell, 2013, p 221–246), that is, changes that can be systematically measured and are backed by theoretical analysis (Hammarström *et al.* 2019). For instance, euphemistic shifts often depend on social norms. When the word 'weed' acquired the sense of 'cannabis', this shift was motivated by the intention to obscure the literal meaning, given that cannabis carried negative or prohibited connotations within that particular cultural and historical context. Capturing this type of change requires not only knowledge of the lexical relation between 'grass' and 'cannabis', but also an understanding of external social and cultural factors.

Another example concerns cases in which a word's sense becomes more general or more specific. Such changes cannot always be detected through direct comparison of the two senses, as the more general sense may still appear appropriate in both contexts, thus obscuring the shift. We focus on three factors of semantic change: **orientation** (i.e., whether a word's meaning becomes more pejorative or positive, e.g.,, the word awful acquired a negative meaning), **relation** (i.e., whether a word's meaning changes through figurative usage, 'head' used to address the director of a company), and **dimension** (i.e., whether a word's meaning becomes broader or more specific, 'cell' used to mean only a room in prison).

These poles of change also encode fundamental aspects of meaning: denotative (dimension), connotative (orientation), and cognitive (relation). Some of these properties, as we detail later, are crucial for understanding the intent of a sentence, for writing tools, to assist with translation, or to interact with users (e.g., in a social robot). Finally, this will open new opportunities to improve the reasoning capabilities of NLP tools (Danilevsky *et al.* 2020).

[a]While NLP resources have hundreds of thousands of examples, datasets for semantic change generally cover fewer than a hundred words Kutuzov *et al.* (2022).

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

To our knowledge, this is the first survey to comprehensively cover semantic change characterization by consolidating and categorizing existing approaches and identifying research gaps. It introduces formal descriptions and proposes enhancements for a more robust understanding of semantic change. By synthesizing prior research and addressing its limitations, this survey aims to streamline existing work and advance our understanding of this complex phenomenon.

The remainder of the paper is structured as follows: Section 2 reviews existing surveys, highlights their differences from this work and presents our survey method. In Section 3, we investigate available corpora for semantic change analysis. In Section 4, we explore methods for representing and characterizing meaning change. In Section 5, we present a formal model derived from the reviewed methodologies. Section 6 presents our analysis and open questions. Finally, Section 7 concludes the survey and outlines future work.

## 2. Lexical semantic change

For the characterization of change, we adopt the mainstream linguistic typology for semantic change (Traugott, 2017): (i) broadening and narrowing, (ii) amelioration and pejoration, and (iii) metonymization and metaphorization. Before studying change characterization, we first review work on change *identification*. This is a well-developed topic that focuses on evaluating whether a word's meaning has changed. Papers addressing change identification often state the need for further characterization of the type of change, but the state of computational change characterization remains unclear. This conclusion motivated the work presented in this paper and differentiates our survey from existing ones on LSC.

In the following subsections, we present existing surveys on the semantic change problem and identify gaps in the current literature. Inspired by their definitions, we propose our own typology of changes. The last subsection introduces the objectives of this survey and our plan to address the gap in semantic change characterization.

### *2.1 Related works on identification of lexical semantic changes*

Existing surveys on semantic change regularly cover detection but seldom focus on characterization, which is the topic of this paper. In the context of identification, Tang (2018) proposes splitting LSC analysis into four steps: obtaining and preparing a diachronic corpus, identifying word senses, modeling semantic change, and presenting and validating hypotheses. This empirical, rather than historical, paradigm enables the study of semantic change through scientific and statistical methods, allowing findings to be reproduced and tested – a key element for computational research.

Tang's framework structures the survey by describing studies related to each step. Change identification is mainly discussed in the 'modeling semantic change' section, which explains the difference between historiographical and empirical analysis. However, it does not cover approaches that classify changes into categories like broadening, narrowing, pejoration, amelioration, or figurative shifts.

A second survey on computational approaches to LSC (Tahmasebi *et al.* 2018) focuses on word-level differentiation, sense differentiation, and lexical replacement. While it includes some types of change categories (e.g., metaphor, metonymy), the main change categories discussed are broadening, narrowing, and correlated types (split, join, birth).

However, one recommendation from the authors is to conduct deeper studies on methods for automatically detecting other types of change and using them to build word histories. The authors provide an in-depth analysis of the linguistic literature, grounding each domain, followed by evaluation strategies and applications. The difficulty of comparing approaches is highlighted and attributed to the absence of standard metrics and reference datasets, but also to the inherent challenges of the task itself.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

Kutuzov *et al.* (2018) survey word representation models and data for semantic change identification and describe a framework for comparing these methodologies. The authors focus on defining consistent terminology and coherent evaluation practices. In this work, they claim there's still a gap in studying sub-classifications of change:

> 'However, more detailed analysis of the nature of the shift is needed. [. . .] This problem was to some degree addressed by Mitra *et al.* (2015), but much more work is certainly required to empirically test classification schemes proposed in much of the theoretical work [. . .]'

For the study of LSC and NLP in general, there are four main ways to represent word meaning: (i) frequency of co-occurrence, (ii) graphs of associated words or meanings, (iii) topic modeling for a list of meaningful terms, and (iv) word embeddings. Initial studies in the field employed graph- or frequency-based methods. Later approaches benefited from advances in the state of the art, replacing discrete methods with those based on static and contextual embeddings. Although contextual embeddings represented a massive improvement for a series of NLP tasks (Devlin *et al.* 2019), the improvement for LSC did not follow this trend (Schlechtweg *et al.* 2020), showing that there is still room for improvement in this domain (Zhou, Ethayarajh, and Jurafsky 2021; Kutuzov, Velldal, and Øvrelid 2022).

Other surveys in LSC investigate the phenomenon of change at the ontological level, trying to discover which laws govern the semantic evolution of words. For example, which kinds of words are more prone to broadening, given their current properties? In this context, Dubossarsky *et al.* (2017) investigate existing evidence for claims such as (i) word frequency affects meaning change, (ii) a negative correlation exists between meaning change and prototypicality, and (iii) a positive correlation exists between meaning change and polysemy. The authors observed that laws proposed in previous studies are affected by model or data bias. When other investigators reduced the bias, the 'laws' either did not hold or were less precise than initially reported.

Research in this field also delves into methodological issues, including the challenges of characterizing semantic shifts. As pointed out by Hengchen *et al.* (2021), few studies detail the nature of these changes, mainly because it is difficult to model meaning from textual data. The paper further concludes that no particular measures for quantifying category change exist and that a consensus on a taxonomy for classifying types of change is required.

In the work of Montanelli and Periti (2023), the authors review methodologies based on contextualized word embeddings. They categorize methodologies for training contextualized embeddings based on how they represent meaning, how time information is presented to the model, and what kind of learning procedure is performed. In this paper, the authors also discuss the importance of characterization:

> 'Most of the literature papers do not investigate the nature of the detected shifts, meaning that they do not classify the semantic shifts according to the existing linguistic theory (e.g., amelioration, pejoration, broadening, narrowing, metaphorisation, and metonymisation) [. . .]. These studies could be crucial to detect "laws" of semantic shift that describe the condition under which the meanings of words are prone to change. [. . .]'

For this survey, we argue that semantic change characterization needs appropriate methods and tools to analyze each fine-grained category. We also defend that observing variations in a single property cannot describe all forms of LSC well. For instance, changes in the angle between word vectors (Dubossarsky *et al.* 2017) are not the best solution to identify the narrowing or broadening of a sense (Tahmasebi *et al.* 2018). Specific methods for identifying different types of change could lead to a better understanding of LSC than straightforward dissimilarity measures.

This survey addresses the need to depict the characterization problem in semantic change and present recent works dealing with this problem. In the next section, we define the survey objectives and the criteria for selecting papers on change characterization.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

### *2.2 Objectives*

Different terms refer to LSC, such as semantic drift or semantic shift. In this paper, we distinguish between the identification of a change and the characterization of that change. The former refers to detecting that a word's meaning has changed, while the latter refers to describing how it has changed. We formally differentiate these problems in Subsection 5.1. The analysis of work on the identification problem has been well covered by Tahmasebi *et al.* (2018), Kutuzov *et al.* (2018), and Tang (2018).

Semantic change occurs from the need to convey more complex concepts and ideas with a restricted set of known words (Hock and Joseph 2019). While some authors categorize changes by their cause (e.g., social, cultural), others use linguistic mechanisms (e.g., metaphor, metonymy) as the basis for classification. The focus of our literature review is on the characterization of change, where we look for common properties that allow for grouping identified semantic changes into categories. From a computational perspective, determining the cause of a semantic change is very complex to implement, and, to the best of our knowledge, no computational approach has been implemented yet.

Frameworks have been proposed on how to track LSC, such as the Cambridge and McTaggart views (Tang *et al.* 2013), which did not initially consider computational methods. These approaches differ in how they use temporal information to identify changes: while the Cambridge view involves comparing two static corpora from different time periods, the McTaggart view involves tracking the evolution of a word's usage over time. The Cambridge approach best suits the computational perspective, as the McTaggart approach would require historiographical work to acquire socio-cultural knowledge that goes beyond contextual analysis of word occurrences (Tang *et al.* 2013, p 4–7).

This core distinction causes considerable disagreement, as empirical and historiographical methods are based on different ideas and focus on distinct aspects of research. Empirical methods rely on direct observation and measurable data, whereas historiographical methods involve interpreting past events and written records. Since these two approaches operate in different ways and aim to answer distinct kinds of questions, they often yield contrasting views.

For this review, we assume the Cambridge setting for change analysis, using linguistic figures as a basis for change categories, which aligns with current state-of-the-art methods for computational LSC (Pivovarova and Kutuzov 2021). Given these considerations, we adopt the typology from Traugott (2017), which is also predominant in the literature, as seen in Koch (2016, p 21–66), Hock and Joseph (2019, p 189–222), and Campbell (2013, p 221–246):

- **Broadening:** A word gains new meanings, making it applicable to more concepts, for example, 'cloud' as a computing infrastructure.
- **Narrowing:** A word's meaning becomes more restricted, applying to fewer concepts. For example, 'gay' once meant 'jolly' or 'festive' but now primarily refers to homosexuality.
- **Amelioration:** A word gains a more positive sense. For example, *nice* changed from 'foolish, innocent' to 'pleasant.'
- **Pejoration:** A word is used with a worse connotation. For example, Old English *stincan* ('to smell sweet or bad') changed to *stink*.
- **Metonymization:** A change based on contiguity or association between concepts, for example, *board* from 'table' to 'a governing body of people who sit at a table'.
- **Metaphorization:** Conceptualizing one thing in terms of another, for example, using the *heart* to represent 'feelings.'

Since the types in this typology form natural pairs that often involve similar computational challenges (e.g., broadening/narrowing requires sense counting; amelioration/pejoration requires sentiment analysis; metaphorization/metonymization requires a relatedness measure

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

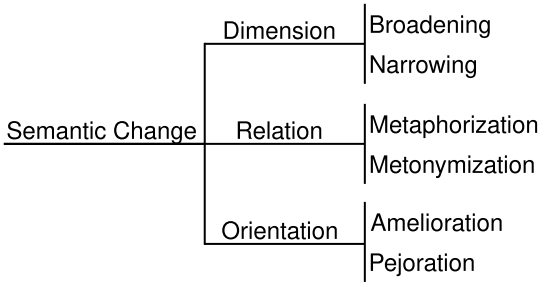


**Figure 1.** Taxonomy for the poles of Lexical Semantic Change, based on the work of Blank *et al.* (2003); Koch (2016); Hock and Joseph (2019).

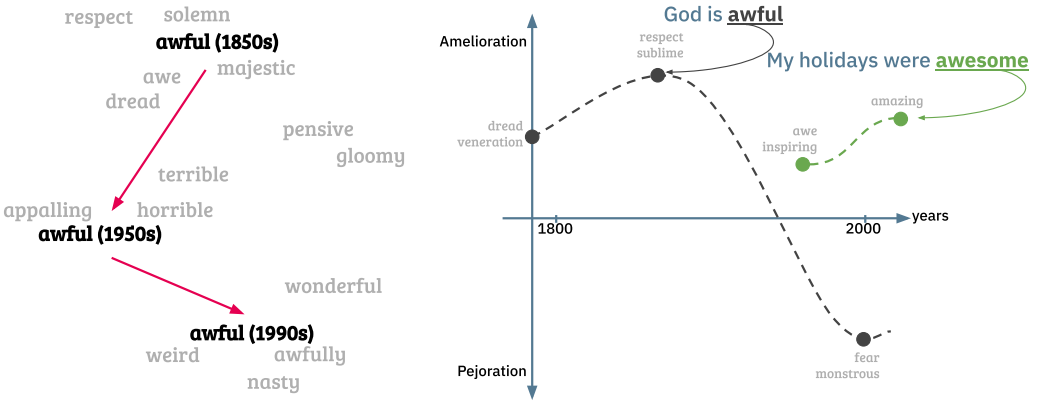


**Figure 2.** Change in meaning and orientation for the word awful. On the left side, we reproduce the figure from Hamilton *et al.* (2016b) that shows the evolution of the word 'awful' in the embedding space. On the right side, we present the hypothetical function (f) for this word over time.

(Blank, 2003; Koch, 2016; Hock and Joseph 2019)), we group them into three overarching poles to better compare and categorize research. We only include approaches that compare corpora, as this typology is designed for diachronic studies rather than synchronic variation (e.g., across domains). Figure 1 illustrates the hierarchical taxonomy for LSC.

We assign studies that track sense broadening or narrowing to the **dimension** ($\mathscr{D}$) pole, where words gaining or losing senses see a change in their dimension value (see Subsection 5.2). This pole is self-complementary, as the change in the dimension's value addresses both broadening and narrowing. A change in dimension could signal, for example, that a word requires more careful translation (Mohammed, 2009).

We group changes related to metaphorization or metonymization into the **relation** ($\mathscr{R}$) pole. These changes occur when a new sense develops from an existing one through a figurative link (see Subsection 5.3). Relations like hypernymy, hyponymy, and synonymy are also factors for semantic change (Koch, 2016; Hock and Joseph 2019) as they involve a conceptual relation and can potentially be encompassed in this pole.

For the **orientation** ($\mathscr{O}$) pole, we group studies that identify amelioration, pejoration, or any perceptual change in the sentiment associated with a word (e.g., criticism, compliment). This pole also encompasses many forms of change in perception, as some studies suggest that concepts like 'taboo' also affect changes in word usage (Koch, 2016).

To illustrate a word that changed in orientation, consider 'awful'. In the 1850s, it meant 'full of awe', but today it means 'extremely disagreeable'. As the first sense is semantically positive and the second is negative, this word underwent pejoration, as illustrated in Figure 2. Notice that the

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Table 1.** Search terms used to discover articles related to change characterization

| Characterization | Search terms |
|---|---|
| Dimension | "semantic change broadening", "semantic change narrowing", "novel sense change", "innovative meaning", "new sense identification", "neologism identification" |
| Relation | "metonymy semantic change", "metonymization semantic change", "metaphor semantic change", "metaphorization semantic change" |
| Orientation | "semantic change polarity", "semantic change feeling", "semantic change orientation", "semantic change emotions", "semantic change sentiment" |

word's orientation may have first undergone amelioration before pejoration. In the same figure, the word 'awesome,' which has a close relation to 'awe', gained a more positive sense, undergoing amelioration over time.

In the next section, we describe our method for reviewing the literature based on the poles presented above.

### *2.3 Review criteria*

In this survey, we select papers that represent and analyze words from one of the poles mentioned above in an automatic manner. We exclude works that involve manual analysis, as we focus on computational methods. For the relation and orientation poles, these searches returned many unrelated results, such as studies on synchronic metaphor detection, which do not infer diachronic change and were therefore excluded.

Querying scientific databases required an iterative process to include all relevant terms. In our initial efforts, we searched for '[type of change, e.g., broadening] semantic change' in Scopus, which returned 33 articles; however, none of them were relevant. Due to the lack of consensus in terminology and the rapid pace of the field, we adopted a semantic search approach, starting with key papers and iteratively expanding our search query. Table 1 presents the final set of search terms used to collect articles.

From these initial queries, we could find 11 seed papers that satisfy our criteria. We then employed a network semantic citation graph to recover more papers, Research Rabbit.[b] From the 11 initial papers, the graph suggested 83 new papers; from this, we were able to recover 7 that matched our criteria. We again included these additional 7 papers as seed, obtaining 129 papers. From 129 papers, we obtained 5 more relevant papers. In the last interaction, we obtained 151 paper suggested, but no new relevant papers were found. In Figure 3, we present the graph of documents, where green nodes represent the input papers, and blue nodes represent papers suggested by the tool.

In the next sections, we investigate three elements of semantic change characterization in depth: (i) the corpora ($T$) analyzed, (ii) the representation method ($\mathfrak{R}$) for capturing characteristics of change, and (iii) the kind of characterization ($\{\mathscr{D}, \mathscr{R}, \mathscr{O}\}$) addressed.

## 3. Corpora for semantic change analysis

Corpora for semantic change can be selected based on differences in time (diachronic studies), domain (academic fields (Schlechtweg *et al.* 2019)), culture (social media (Carpuat *et al.* 2013; Shoemark *et al.* 2019), countries), or dialect (Takamura, Nagata, and Kawasaki 2017; Uban, Ciobanu, and Dinu 2021), provided the corpora share a sufficient vocabulary for comparison.

[b] https://researchrabbitapp.com/

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

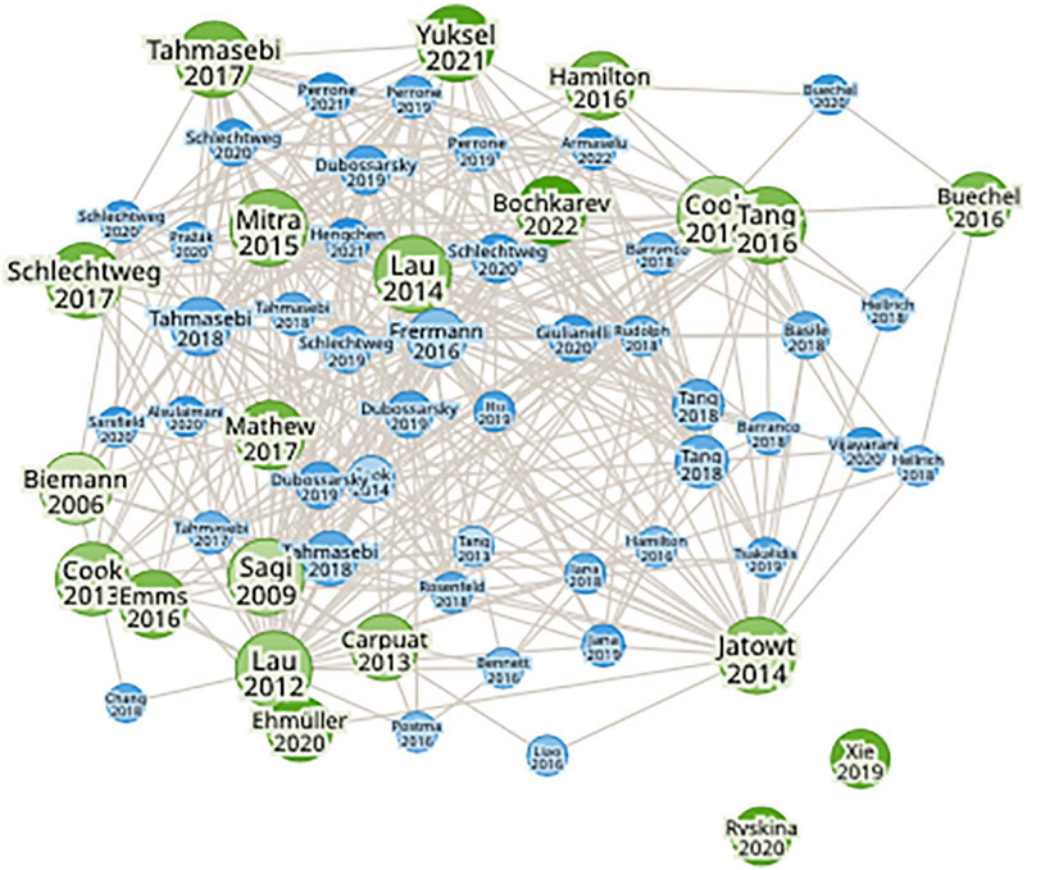


**Figure 3.** Graph of selected works (green) and related articles (blue).

In diachronic studies, some methods divide a corpus into discrete time windows, striking a balance between the size of the window and the quantity of data it contains. In contrast, others use a continuous representation of time, providing the timestamp of each text directly to the model (Rosenfeld and Erk 2018; Hu, Li, and Liang 2019; Wang, Buccio, and Melucci 2021). These corpora were designed for specific purposes. For example, corpora with discrete time bins enable a comparative perspective on meaning, while continuous time representations allow for an evolutionary view of the changes.

With respect to characterization, the collected papers show an increasing interest in diachronic corpora, with many proposing handcrafted rules for automatically annotating change categories (Tang *et al.* 2013). Few works rely on human annotation (Schlechtweg *et al.* 2017). While these approaches have advantages and drawbacks, the data required to train large representation models is often insufficient from a computational perspective.

Most of the works reviewed here follow a specific formula for diachronic studies, one historical corpus ($C_1$) and a more modern corpus ($C_2$). Splitting the Corpus of Historical American English (COHA) Davies (2012), a diversified corpus with a span from the 1810s to the 2000s, into two corpora is the standard choice. Recent state-of-the-art NLP models require training on billions or even trillions of tokens (He, Gao, and Chen 2021), but diachronic corpora are often on the scale of millions of tokens (Schlechtweg *et al.* 2020). Given that increasing the amount of historical training data is often infeasible, a viable approach is to develop better algorithms for learning from limited data. In Table 2 we present the table of diachronic corpus in the reviewed works.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Table 2.** Overview of Various Corpora for Diachronic Studies

| Corpus Name | Description | Statistics (Date, Size) |
|---|---|---|
| CLMET (2025) | A collection of diachronic English corpora covering various text types like literary works, letters treatise and others. | 34 million words, spanning from 1710-1920, divided into three 70-year sub-periods. |
| COHA (2025) | A balanced diachronic corpus of American English compiled from diverse document genres. | Over 400 million words from about 107,000 documents published from the 1810s to the 2000s. |
| People's Daily Newspaper (2025) | A Chinese newspaper corpus used for its faithful recording of contemporary language and events. | Spans from 1946 to 2004, with a yearly average of 10,886,017 tokens. |
| Google N-Gram Book (2025) | The largest available historical corpus, compiled from about 4% of all ever-published books. | Over 1TB of text data, containing about 0.3 trillion words and spanning from 1600 to 2009. |
| Kubhist2 (2025) | A sample of a Swedish historical newspaper archive. | From 1790–1830 containing about 71 million tokens. |
| DTA (2025) | A high-quality collection of historical German texts selected for representativeness and genre balance. | 1,022 texts from the period 1741–1900. |
| Lampeter Corpus (2025) | A collection of British English texts, encompassing various domains such as religion, politics, and law. | Approximately one million words from 1640–1740. |
| BNC (2025) | A collection of primarily written British English sources. | One hundred million words, from the late 20th century. |
| NYT Archives (2025) | A collection of newspaper articles from The New York Times archives. | 13 million articles (1851-present). |
| Helsinki Corpus (2025) | Comprises texts spanning Old English, Middle English, and Early Modern English. The analysis focused on the Middle and Modern English. | Approximately 1.1 million words. Old (1150), Middle (1150–1500), and Modern (1500–1710) |
| Project Gutenberg (2025) | The bulk of English language literary works available through the project's website. | 4034 separate documents consisting of over 290 million words. |

Even with progress in low-resource learning fields like rapid adaptation (Neubig and Hu 2018), few-shot learning (Liu *et al.* 2022), or in-context learning (Min *et al.* 2022), there is still a significant lack of freely available corpora for semantic change studies (Kutuzov *et al.* 2018).

## 4. Methods for semantic change characterization

While interest in detecting semantic change is growing, less effort is being applied to characterizing these changes. Computational linguistics has not yet fully explored the utility of having detailed information about the type of change. For example, a change in relation might have a low impact on a sentiment analysis algorithm, whereas a change in orientation is more likely to affect it.

In this context, meaning representation ($\mathfrak{R}$) methods are highly relevant to understanding what kind of change can be described and characterized. This section investigates representation methods from the literature based on word frequency, topics, graphs, and embeddings. We analyze how each method was used to characterize LSC within our three proposed poles.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

### *4.1 Word frequency*

Methods based on word frequency were among the first computational attempts to tackle semantic change and remain central to many linguistic approaches. Analyzing word frequency offers a direct interpretation of change, which, for characterization, usually involves counting co-occurrences with other words.

#### *4.1.1 Dimension*

To detect changes along the dimension axis ($\mathscr{D}$), Sagi *et al.* (2009) proposed counting co-occurrences of words and phonemes and measuring the cluster density of the cosine distance on a decomposition of the term-frequency matrix. They assumed that Latent Semantic Analysis is appropriate for examining vectors representing the context of each term's occurrence and estimated cohesiveness by measuring the density of these context vectors. While they explored the methodology for taxonomic change analysis, they state that the method in fact measures word polysemy.

Tang *et al.* (2013) proposed a framework for semantic change computation based on the assumption that it is an inherently successive and diachronic process. They distinguished between the word level (overall meaning) and the sense level (individual senses), building their approach around these two notions. A key step involves first identifying a word's senses in a corpus and then analyzing their diachronic distribution.

Emms and Jayapal (2016) proposed a generative conditional model to detect if a word has acquired a new sense (a semantic neologism). They modeled the word's sense distribution using Gibbs sampling, applying the model to time-stamped text to infer how the probabilities of a word's senses change over time. The authors obtain all usages for a word in a specific year ($C_1$) from Google N-gram Corpus and compare with the usage frequency for the following year($C_2$), obtaining n-grams that change above a threshold.

Bochkarev *et al.* (2022) used a neural network to predict whether a word is used as a named entity. With this approach, they constructed a temporal view of word usage by tracking occurrences as nouns versus non-nouns over time, aiming to detect if the word acquired a new meaning.

#### *4.1.2 Relation*

The work of Tang *et al.* (2013) transforms the problem of relation identification into detecting change patterns in time series data. They analyze semantic change at the word and individual-sense levels, using rules to infer whether a word is used metonymically or metaphorically. A different approach computes the entropy of target words to identify if a word was used in a non-standard context (Schlechtweg *et al.* 2017). The authors split the usage of words from the Deutsches Textarchiv (DTA) in two time periods (varying by word) and measure the increase in entropy as a proxy for metaphorization. According to the authors, a word's higher entropy makes it more likely to be used in a newly created metaphorical sense. However, the authors do not differentiate this from the emergence of a new, unrelated sense. They use a corpus of German documents to validate their approach.

#### *4.1.3 Orientation*

Frequency metrics that detect co-occurrences with sets of 'good' or 'bad' sentiment words have been widely used to infer the amelioration and pejoration of a word over time, especially for neologisms (Cook and Stevenson 2010; Jatowt and Duh 2014). Similarly, Buechel *et al.* (2016) used seed words in German and English to track emotions captured in corpora. They represent these emotions in the Valence-Arousal-Dominance framework rather than simple positive/negative connotation. A limitation of the latter study is that it doesn't track the evolution of orientation in words.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

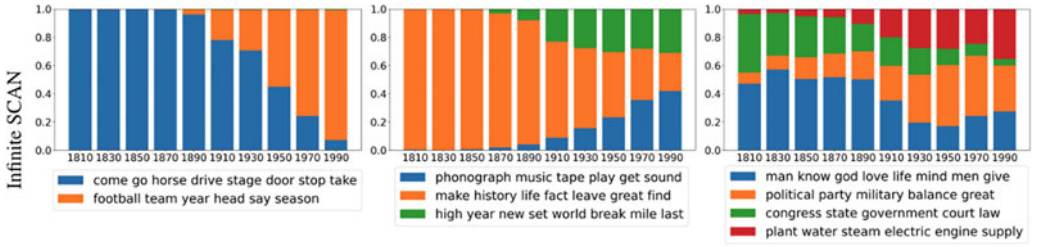


**Figure 4.** Figure adapted from Inoue *et al.* (2022). The stacked bar plots represent the topics obtained over time for the words 'coach', 'record', and 'power' respectively. We can observe new senses emerging and becoming dominant.

### *4.2 Topics*

Topic modeling is a statistical method used to identify the main themes in a collection of documents by grouping words that frequently appear together. For example, a topic model might identify the topic of 'banking' in a news corpus by grouping words like 'government', 'money', and 'insurance'.

Topic modeling can be used to understand the meaning of words and phrases by tracking the different topics a word is associated with in various contexts (e.g., time, domain, culture). For example, a topic model could be used to understand the different meanings of 'bank' in a financial context versus a geographical one. The different topics associated with a word are tracked and compared to identify and characterize changes in its senses.

#### *4.2.1 Dimension*

These approaches assume that changes in the number of topics assigned to a word can indicate changes in dimension. Lau *et al.* (2012) introduced the idea of novelty as the proportion of a word's usages corresponding to senses in a focus corpus versus a reference corpus. The goal is to use word sense induction based on topic modeling to detect if a word is used with a new meaning.

Cook *et al.* (2013) extended this work by quantifying a topic's usage, estimating that a previously 'culturally silent' topic that becomes frequent in a focus corpus is highly likely to contain neologisms. They introduced the concept of topic relevance to compute a ranked list of potential neologisms.

Lau *et al.* (2014) improved the definition of novelty by adding the word's frequency in the corpus, then proposed combining novelty and relevance to rank neologisms. Frermann and Lapata (2016) introduced SCAN, a Bayesian topic model where topics are preserved across adjacent time periods. Inoue *et al.* (2022) introduced InfiniteSCAN, an improvement over SCAN. While most topic models rely on a predefined number of topics, InfiniteSCAN uses a logistic stick-breaking process to adapt the number of topics dynamically. This modification allows it to track a varying number of senses, enabling it to count each meaning to measure broadening and narrowing, as shown in Figure 4. The authors split Clean COHA into intervals of 20 years to represent the senses.

#### *4.2.2 Relation*

To the best of our knowledge, no methodology for tackling conceptual change with topic-based representation has been published, although a work uses it in a synchronic detection setting (Heintz *et al.* 2013).

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

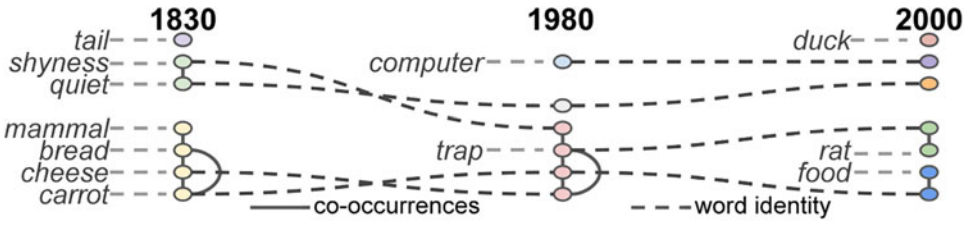


**Figure 5.** Adaptation from Ehmüller *et al.* (2020). Ego-network, built from the word co-occurrence graph, for 'mouse'. We observe that in 1830 it was used with the sense of 'weak' and 'rat,' where in 1960 the sense of 'computer device' emerged.

#### *4.2.3 Orientation*

To the best of our knowledge, no methodology approaching orientation change with topic-based representation has been published. Wankhade *et al.* (2022) use topic models in their workflow to refine data for a sentiment classifier, but topics are not directly used to detect a change in feelings.

### ***4.3 Graphs***

Graph structures are a natural way to represent relationships between words. They are often constructed manually from dictionaries or automatically based on word co-occurrence.

Co-occurrence graphs can capture richer contextual information than count-based methods because their edges represent more complex relationships. For example, the words 'feline,' 'cat,' and 'pet' would likely form a clique in a co-occurrence graph, indicating they are often used together and share close meanings.

#### *4.3.1 Dimension*

Comparing the size and structure of a graph over time can provide rich information about a word's semantic change. Biemann (2006) was among the first to use this approach to detect sense broadening. He created a graph of associated nouns from a corpus and then applied the Chinese whispers algorithm to find clusters. By comparing these clusters over time, he could infer whether a word had acquired a new meaning.

Mitra *et al.* (2015) analyzed tweets, counting word co-occurrences in bi-grams to build a graph. They ran a graph-clustering algorithm (Chinese Whispers) to determine a set of senses, where each cluster represents a different sense. By comparing these sets of senses over time, they could infer if a word's senses split or became broader.

Mathew *et al.* (2017) improved the filtering and sense selection methods of previous works (McCarthy *et al.* 2004; Lau *et al.* 2014; Mitra *et al.* 2015) to better capture sense evolution in Google Books and newspaper datasets. Tahmasebi and Risse (2017) considered a graph of two-word noun phrases as a cluster of meaning, then merged sub-graphs based on Lin similarity (Lin, 1998). They tracked the change in dimension by analyzing how these sub-graphs change over time.

Ehmüller *et al.* (2020) built a co-occurrence graph, filtering it with point-wise mutual information to remove non-relevant edges, and then created ego-networks to run graph-clustering algorithms. This approach allowed them to build a sense tree – a tree of word usage evolution where new senses are new branches, as shown in Figure 5. Similar to previous work, the authors restricted their analysis to nouns only.

#### *4.3.2 Relation*

Graphs are a good way to represent word relations, and some approaches have used them to detect metonymy usage, for example, Teraoka (2016); Pedinotti and Lenci (2020). Their models use WordNet or a graph-based distributional model to determine a contextualized representation

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

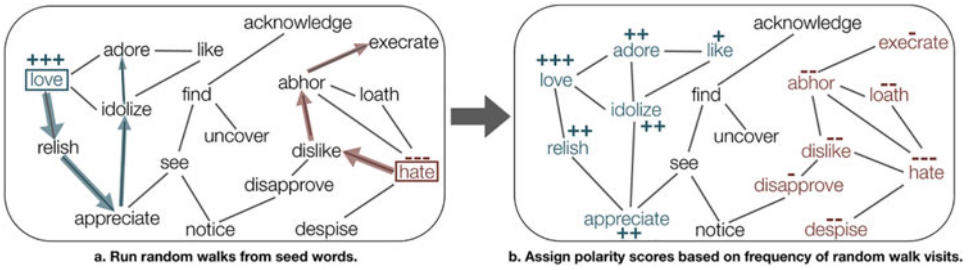


**Figure 6.** Adapted from Hamilton *et al.* (2016a). The SentiProp algorithm propagates the polarity from seed words based on the distance of the connected nodes. Words are connected based on co-occurrence statistics.

of word semantics. However, to the best of our knowledge, no work has proposed using graphs at a representation level to compare which words have changed their relations between corpora.

#### *4.3.3 Orientation*

The orientation of words is generally inferred from similarity or proximity to seed words. Many word graphs are constructed for this purpose, and some works (Reagan *et al.* 2017; Barnes *et al.* 2021) use these graphs to infer orientation. Hamilton *et al.* (2016a) used the SentiProp algorithm to propagate a word's polarity to its synonyms and connected words, analyzing changes in polarity for words across different times and domains. In Figure 6, we reproduce the illustration of this algorithm.

### ***4.4 Embeddings***

Word embeddings are representations of words as numeric vectors. The vectors are constructed such that words with close values in the vector space are expected to have similar meanings. The distance between these vectors can help in interpreting their sense.

Word embeddings have been used for various tasks, including LSC identification and characterization. This is done by comparing word vectors between two moments in time; the distance between them, and relative to neighboring vectors, can indicate if the word's sense has changed (Hu *et al.* 2019).

#### *4.4.1 Dimension*

Ryskina *et al.* (2020) proposed a statistical approach to identify neologisms based on frequency count and semantic sparsity (measured with embeddings). They claim that words and concepts are subject to supply and demand, so emerging concepts (i.e., neologisms or sense broadening) appear to populate sparse semantic regions until the space becomes uniformly covered.

In Giulianelli *et al.* (2020), the authors use BERT to create contextualized embeddings of words and then cluster them to obtain usage types. They apply entropy difference, Jensen-Shannon divergence (JSD), and average pairwise distance (APD) as proxies to measure word broadening and narrowing, finding that higher APD and JSD were the best metrics for broadening.

By analyzing the deviation of a Gaussian word embedding, Yüksel *et al.* (2021) identified a way to measure narrowing and broadening in word senses. In Moss (2020), the authors also performed per-sense training of these embeddings. In Figure 7, we observe the distribution of the word 'gay' becoming wider and closer to 'homosexual'.

Lautenschlager *et al.* (2024) use XL-LEXEME (Cassotti *et al.* 2023) clustered embeddings to progressively discover new senses in the historical COHA corpus. They frame the task as Unknown Sense Detection: given an initial dictionary, the goal is to identify unrecorded senses.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

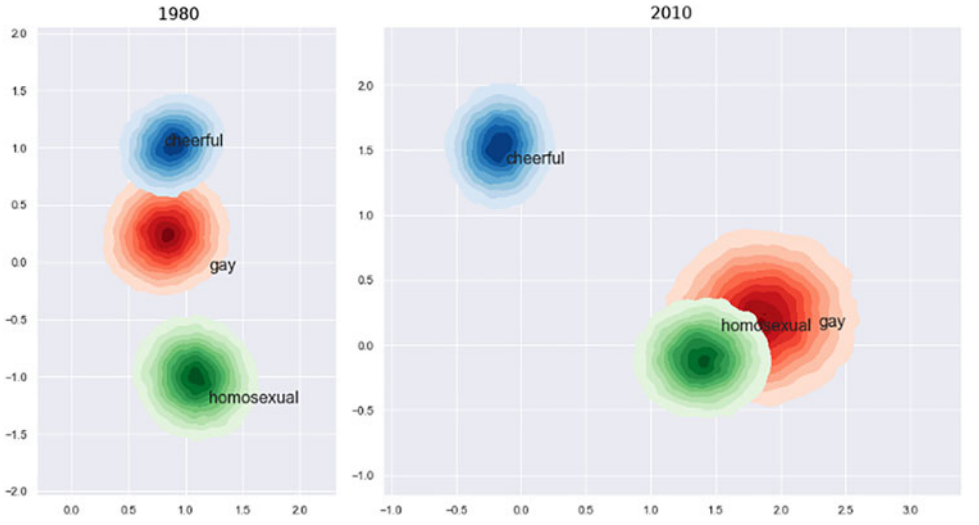


**Figure 7.** Plot obtained from Moss (2020). The author represents words as Gaussian embeddings, and analyzes it's variance and proximity over time. The word gay increased in variance and got closer to 'homosexual'.

The system compares a word's usage to all its dictionary senses, and if the highest similarity score falls below a threshold, the usage is flagged as a potential new sense.

#### *4.4.2 Relation*

Fonteyn and Manjavacas (2021) investigate the metaphorization of the concept 'to death'. They combined cluster analysis and sentiment analysis to contextualize occurrences of this phrasal verb and detect if metaphorization had taken place. Another approach is to create a training corpus and use it to train a BERT model to predict if a word is metaphorical or literal. The work of Maudslay and Teufel (2022) follows this approach and proposes a framework to detect words that started being used metaphorically.

#### *4.4.3 Orientation*

Methods based on word embeddings have been widely used for many sentiment analysis tasks (Raffel *et al.* 2019) and to investigate sentiment at the lexicon level (Ito *et al.* 2020). In the context of LSC characterization, Fonteyn and Manjavacas (2021) used the distance between the word vectors of 'good' and 'bad' to measure the polarity of the expression 'to death,' as illustrated in Figure 8.

Similarly, Hellrich *et al.* (2018) and Xie *et al.* (2019) employed diachronic static embeddings to analyze the sentiment of words over time. Hellrich et al. train word embeddings in COHA from 1810-1830 as $C_1$ and a VAD lexicon Warriner *et al.* (2013) as $\mathfrak{R}(C_2)$, tracking for orientation of words the valence-arousal-dominance framework. They access the polarity of the word based on the k-nearest-neighbors from a VAD sentiment lexicon. Xie et al. use 'moral' seed words to compare, like 'cheating' and 'fairness,' also for COHA, as shown in Figure 9.

### *4.5 Synthesis and summary*

In this section, we synthesize the pros and cons of representation methods for characterizing semantic change and present a table that provides an overview of the current poles of change addressed in the literature.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

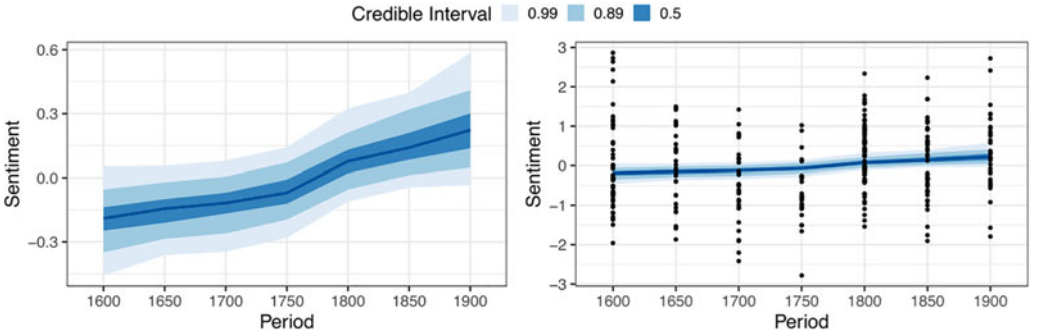


**Figure 8.** Adaptation from Fonteyn and Manjavacas (2021). The line plot shows the evolution of the polarity of the multiword 'to death.' It went from a negative concept to a more positive one, ameliorating the dominant sense.

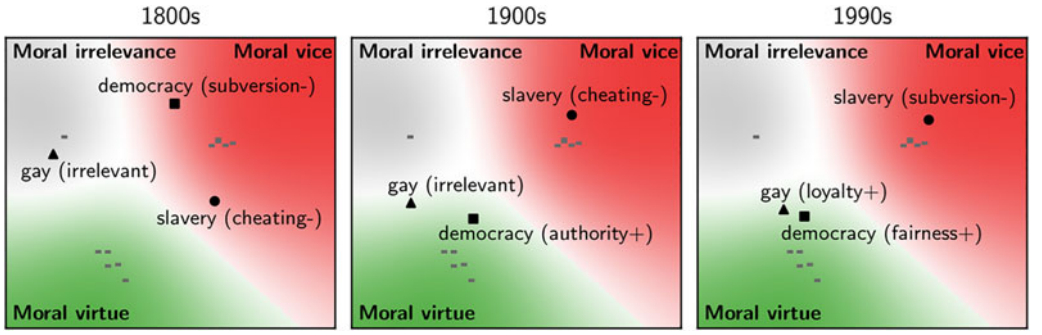


**Figure 9.** Illustration adapted from Xie *et al.* (2019). The figure shows how moral sentiment toward slavery, democracy, and gay rights evolved over two centuries, mapped in a 2D embedding space.

**Word frequency**-based methods are simple to implement, offer an easily interpretable analysis, and require a relatively small corpus of data. However, these methods struggle with capturing word similarity (Mikolov *et al.* 2013) and compositionality (Mitchell and Lapata 2010), limiting their semantic analysis capabilities. While N-grams slightly improve compositionality, they are constrained by Zipf's law (Zipf, 1949), requiring an exponential increase in data to increase the size of N. Additionally, frequency alone cannot distinguish between sense broadening and narrowing, requiring extra steps to handle polysemy. Co-occurrence tables also fail to correlate well with polysemy or semantic shifts. In parallel studies, these methods perform poorly in detecting word orientation (Socher *et al.* 2013) and struggle with contexts involving sarcasm, irony, or negation due to their limited contextual awareness. **Topic modeling** is a powerful text analysis tool with many good libraries implementing state-of-the-art algorithms, and it doesn't require supervision or data annotation. The downside is that these methods require manual interpretation, as statistically derived word clusters can be ambiguous. Automated topic selection using dictionary-based filtering aids in neologism detection but remains insufficient for change characterization. Metrics like coherence and diversity, along with visualization techniques, help reduce manual effort.

**Graph**-based approaches rely heavily on annotated and structured data, which is often costly or, in some domains, unfeasible. However, Large Language Models (LLMs) enable automatic annotation and structuring of information. We have yet to fully leverage modern structures like knowledge graphs for LSC detection. Resources like Wikidata could help model semantic shifts (Giulianelli *et al.* 2023; Tang *et al.* 2023), such as metonymization. Existing lexical databases (WordNet, BabelNet, ConceptNet) assist in change detection but lack temporal information, limiting their effectiveness.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Word embeddings** often exhibit instability, where minor corpus variations alter vector representations, especially for low-frequency or polysemous words (Hellrich and Hahn 2016; Gladkova and Drozd 2016; Günther *et al.* 2019). Solutions such as larger embeddings, quantization, and alignment help mitigate this instability (Hamilton, Leskovec, and Jurafsky 2016b); however, the meaning conflation deficiency (Camacho-Collados and Pilehvar 2018) remains an issue, as a single vector – even when contextualized – does not provide enough information to differentiate fine-grained senses (Ballout *et al.* 2024; Periti *et al.* 2024). Subsequent work has explored various modifications to transformer architectures to enable better representation of meaning (Zhang, van de Meent, and Wallace 2021; Tay *et al.* 2022), but these approaches remain underexplored in the semantic change literature.

Contextualized embeddings (e.g., from ELMo, BERT) enhance sense detection but can produce overly contextualized vectors, resulting in unreliable diachronic analysis and false positives (Kutuzov *et al.* 2022). Recent methods using contrastive learning and similarity-based training enhance isotropy and sense differentiation (Reimers and Gurevych 2019; Gao, Yao, and Chen 2021). Other approaches, like density-based embeddings (e.g., word2gauss), aim to model polysemy directly. The current state of the art in LSC identification fine-tunes cross-lingual models trained on Word Sense Disambiguation (WSD) tasks, benefiting from transfer learning across languages.

While **generative Artificial Intelligence (AI)**, particularly LLMs, is rapidly transforming NLP, its application to the systematic characterization of semantic change in diachronic data remains largely unexplored. However, some initial studies have successfully leveraged these models for related tasks, such as generating synthetic data for LSC detection (Cassotti, De Pascale, and Tahmasebi 2024) and automatically comparing word senses in context (de Sá *et al.* 2024).

Table 3 summarizes the literature surveyed in the previous sections. The rows list relevant papers, and the columns represent the key properties of LSC evaluated in this survey. An 'X' indicates that the paper applied the sense computation technology (listed in the fifth column) to analyze the corresponding pole.

We can observe that no single methodology characterizes all three poles, highlighting the complexity of the problem and researchers' tendency to focus on a specific type of change. The majority of studies were developed to characterize changes in the dimension pole, which seems to be where computational approaches achieve the best results. Another observation is the underexplored use of graph- and topic-based methods to represent sense, compared to embedding- and frequency-based methods. It is also important to note that the most recent works use embeddings, showing a current trend toward this type of approach. In Table 4 we summarize these findings.

Although computational linguistics has been used to characterize LSC for over a decade, there is still no standardized formalization of these changes. Each author proposes their own solution, and many interpretations of change are presented in a textual or semi-formal format. To contribute to this discussion, we propose a set of formal descriptions of change characterization in the next section.

## 5. Formalization

Formalizing a problem is a foundational step in computer science, enabling computational reasoning and the design of algorithms. A shared formalization ensures that researchers operate with consistent assumptions and terminology, facilitating meaningful comparison across methodologies. In our review, we observed that although many studies conceptually addressed the same problem, the absence of a unified formal framework hindered direct comparison. To bridge this gap, we propose a formalization grounded in the surveyed literature and informed by linguistic theory, focusing specifically on semasiological comparative analysis; that is, comparing word meanings across time without invoking external socio-cultural knowledge.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Table 3.** Comparison table between studies highlighting the type of methods utilized and category of change. We mark which kind of characterization the work conducts $\mathscr{D}$ for dimension, $\mathscr{R}$ for relation, and $\mathscr{O}$ for orientation. Also, we highlight the main representation method ($\mathfrak{R}$) for the word meaning

| Methodology | $\mathscr{D}$ | $\mathscr{R}$ | $\mathscr{O}$ | $\mathfrak{R}$ |
|---|---|---|---|---|
| Biemann (2006) | X | - | - | Graph |
| Sagi *et al.* (2009) | X | - | - | Frequency |
| Lau *et al.* (2012) | X | - | - | Topics |
| Cook *et al.* (2013) | X | - | - | Topics |
| Tang *et al.* (2013) | X | X | - | Frequency |
| Lau *et al.* (2014) | X | - | - | Topics |
| Mitra *et al.* (2015) | X | - | - | Graph |
| Emms and Jayapal (2016) | X | - | - | Frequency |
| Mathew *et al.* (2017) | X | - | - | Graph and Topics |
| Tahmasebi and Risse (2017) | X | - | - | Graph |
| Ryskina *et al.* (2020) | X | - | - | Embeddings |
| Giulianelli *et al.* (2020) | X | - | - | Embeddings |
| Moss (2020) | X | - | - | Embeddings |
| Ehmüller *et al.* (2020) | X | - | - | Graph |
| Yüksel *et al.* (2021) | X | - | - | Embeddings |
| Inoue *et al.* (2022) | X | - | - | Topics |
| Bochkarev *et al.* (2022) | X | - | - | Frequency |
| Lautenschlager *et al.* (2024) | X | - | - | Embeddings |
| Schlechtweg *et al.* (2017) | - | X | - | Frequency |
| Fonteyn and Manjavacas (2021) | - | X | X | Embeddings |
| Maudslay and Teufel (2022) | - | X | - | Embeddings |
| Cook and Stevenson (2010) | - | - | X | Frequency |
| Jatowt and Duh (2014) | - | - | X | Frequency |
| Buechel *et al.* (2016) | - | - | X | Frequency |
| Hamilton *et al.* (2016a) | - | - | X | Graph |
| Hellrich *et al.* (2018) | - | - | X | Embeddings |
| Xie *et al.* (2019) | - | - | X | Embeddings |

To the core of this formalization, a central challenge arises in defining what 'meaning' is. While linguistic literature often invokes terms like 'concepts' and 'sentiments' to describe meaning (Blank *et al.* 2003; Koch, 2016; Traugott, 2017), these notions are rarely operationalized in a way suitable for formal modeling. This lack of clarity makes the computational implementation of semantic analysis particularly complex.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Table 4.** Pros and Cons of Representation Methods for Semantic Change Characterization

| Pole of Change | Representation | Pros | Cons |
|---|---|---|---|
| **Dimension** (Broadening/Narrowing) | **Frequency** | - Simple and interpretable.<br>- Good for detecting neologisms (new words). | - Struggles to distinguish distinct senses (polysemy).<br>- Can't differentiate sense gain from higher usage (trend). |
| | **Topics** | - Effective at discovering new senses (sememes).<br>- Unsupervised methods don't require labels. | - Requires manual interpretation to label topics (senses).<br>- Topic granularity can be hard to control. |
| | **Graphs** | - Intuitive for modeling sense splitting and merging.<br>- Can create visual, interpretable sense histories. | - Performance is sensitive to graph construction methods.<br>- Can be computationally intensive. |
| | **Embeddings** | - Can capture subtle semantic shifts.<br>- Density-based methods can model polysemy directly. | - Standard embeddings conflate multiple senses into one vector.<br>- Contextual embeddings can create false positives. |
| **Relation** (Metaphor/Metonymy) | **Frequency** | - Entropy can be a simple proxy for non-literalness. | - Cannot distinguish figurative change from new, unrelated senses.<br>- Highly unreliable and underexplored. |
| | **Topics** | - (Theoretical) Could potentially group figurative usages. | - Largely unexplored for diachronic characterization.<br>- Unclear how to separate from other types of change. |
| | **Graphs** | - Natural fit for modeling relationships between concepts.<br>- Could leverage knowledge graphs (e.g., WordNet). | - Severely under-investigated for this task.<br>- Requires rich, structured relational data, which is rare. |
| | **Embeddings** | - Can be fine-tuned on annotated data to detect metaphors. | - Requires labeled data for supervision.<br>- Hard to interpret *why* a vector represents a figurative use. |
| **Orientation** (Amelioration/Pejoration) | **Frequency** | - Simple to implement via co-occurrence with seed words. | - Easily misled by negation, irony, and sarcasm.<br>- Relies on the stability of seed words over time. |
| | **Topics** | - (Theoretical) Could identify emerging sentiment-laden topics. | - Not used directly for characterization; very little research.<br>- Lacks fine-grained, word-level control. |
| | **Graphs** | - Can propagate sentiment from seed words through relations. | - Performance is highly dependent on graph quality.<br>- Still relies on potentially unstable seed words. |
| | **Embeddings** | - State-of-the-art for most sentiment tasks.<br>- Can measure orientation by projection onto a sentiment axis. | - Can inherit and amplify biases from training data.<br>- Still often relies on the assumption of stable seed words. |

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

To address this, we adopt a minimal form of sense objectivism; the idea that word senses can be treated as identifiable units at specific points in time. Where the senses were conveyed by a community of speakers during the production of the corpus under study. While absolute objectivity is philosophically untenable (Geeraerts, 1997; Wittgenstein, 2009, pg. 182, pg. 32), this abstraction is both methodologically necessary and practically effective (as we target static corpora). It mirrors other linguistic constructs like phonemes: not directly observable or in practical terms precise, yet essential for systematic analysis. Indeed, the very concept of semantic change presupposes a stable referent; without identifiable senses, we cannot speak coherently of change; change compared to what?

This pragmatic stance aligns with both historical linguistic practice (e.g., stating that 'gay' once meant "happy" and now means 'homosexual') and computational models, which require discrete sense representations (e.g., embeddings, clusters) across time slices. Our position is not ontological but operational: sense objectivism enables formal articulation, empirical comparison, and computational modeling of LSC. Without this assumption, both theory and method lose coherence.

We build our formalization by posing the meaning of a word as a set of senses, that is, glosses, inspired by Hock and Joseph (2019, p 194), and define all constructions in set theory. First, we assume that for a set of text corpora $T$, there is a universe of senses $S_T$ that holds all senses for these corpora. Given an alphabet $\alpha$, we define $\mathscr{S}$ as a function that takes a string from the Kleene closure of $\alpha$, where $\alpha^* = V$ (the possible vocabulary), and returns a subset of $S_T$. For instance, given a word $w \in V$ and a corpus $t \in T$, the subset of senses $\mathscr{S}(w, t)$ for the universe of senses ($S_t$) of the word $w$ in this corpus is:

$$\begin{aligned} \mathscr{S}&: V \times T \to \wp(S_t), \\ \mathscr{S}(w, t) &= \{s_1, s_2, \ldots, s_k\}. \end{aligned} \tag{1}$$

We define that two sets of senses are different if, and only if, the symmetrical difference between a set of senses for the word $a$ in a corpus $x$ and the word $b$ in a corpus $y$ is non-empty, where $a, b \in V$ and $x, y \in T$

$$\begin{aligned} \mathscr{S}(a, x) \neq \mathscr{S}(b, y) \iff \\ (\mathscr{S}(a, x) - \mathscr{S}(b, y) \neq \emptyset) \vee (\mathscr{S}(b, y) - \mathscr{S}(a, x) \neq \emptyset), \end{aligned} \tag{2}$$

in simple terms, there are objective senses of the word 'a' in corpus 'x' that do not appear in corpus 'y' for the word 'b,' or vice versa. This underlying assumption is made implicitly in most previous works, even though it is rarely stated explicitly, e.g., when Hamilton *et al.* (2016b) and Inoue *et al.* (2022) state that the word 'record' has a different meaning because it gained the sense of 'tape/music' when comparing the word in corpora before and after 1920.

### *5.1 Semantic change*

Koch (2016, p 23) presents a definition of meaning change that combines semasiological and onomasiological perspectives. In Figure 10, 'belly' was initially associated with the source concept 'bag' (M1). Later, it acquired the meaning of the target concept 'body part between the breast and the thighs' (M2). In computational linguistics, we primarily focus on the semasiological view: our interest lies in the word itself and the evolution of its meanings, rather than the origins of the concepts it represents.

The process of meaning change is depicted on the right side of the figure. At time T1, the word 'belly' possessed only the meaning M1. Subsequently, at T2, it began to be used with the new meaning M2 as well. This usage of M2 increased over time (T3-T5) until it eventually became the dominant meaning (T6). Finally, at T7, the original meaning M1 was completely lost.

To formalize this definition, we say that a word $w$ had a semantic change ($\mathscr{C}$), if the set of senses, given by the function $\mathscr{S}$, evaluated in the corpus $t_1$ is different from the set of senses in $t_2$, given

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

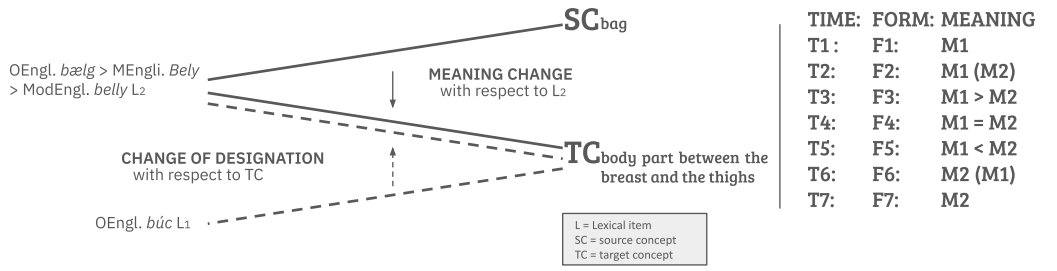


**Figure 10.** Definition of semantic change adapted from Koch (2016, p 23, 25).

that $t_1 \neq t_2$,

$$\mathscr{C}(w, t_1, t_2) = \begin{cases} \text{True} & \mathscr{S}(w, t_1) \neq \mathscr{S}(w, t_2) \\ \text{False} & \text{otherwise.} \end{cases} \tag{3}$$

The subset of words in vocabulary $\bar{W} \subset V$ that suffered a semantic change between corpus $t_1$ and $t_2$ is given by,

$$\bar{W}_{t_1,t_2} = \{w \mid \mathscr{C}(w, t_1, t_2), \quad \forall w \in V\}. \tag{4}$$

These changes in lexical meaning across corpora are caused by many different phenomena, like time period (1930 and 2020, *gay*), culture (Brazil and Portugal, *rapariga*), and domain (engineering and biology, *cell*).

### *5.2 Change in dimension*

We define that a word $w$ had a changed in dimension if the set of meanings increased or decreased, with respect to the cardinality of the set, when comparing two corpus $t_1, t_2$. It means that the word suffered a broadening or a narrowing change between these corpora, that is

$$|\mathscr{S}(w, t_1)| \neq |\mathscr{S}(w, t_2)|, \quad w \in V. \tag{5}$$

Consequentially, if $|\mathscr{S}(w, t_1)| < |\mathscr{S}(w, t_2)|$ it suffered a broadening, and if $|\mathscr{S}(w, t_1)| > |\mathscr{S}(w, t_2)|$, it suffered a narrowing.

This definition is compatible with the concept of innovative/reductive meaning change from Koch (2016, p 24–26), in which words acquire new senses (sememes) by being used with different meanings. It differs, however, from the more general notion of broadening, where words may also lose intension (the precision of their meaning), making them vaguer.

Under this definition, phenomena like sense splitting, birth, and joining are special cases of change in dimension for a single lexical item. For instance, birth or death of a sense is considered a change from or to an empty set. Split and join are instances of broadening or narrowing with respect to a word. For example, the word 'plane' gained the sense of 'aeroplane' after the 19th century, while previously it was only used to describe the act of planning or a flat surface. This new sense greatly increased the dimension of the word 'plane', as many correlated senses, like 'traveling in an aeroplane', were also added.

In this survey, we define broadening as the degree of polysemy in a word. This is aligned with the NLP literature, given its application in data analysis and its measurable nature, compared to the notion of intension. We differ from the definition in Tang *et al.* (2013), where broadening is a superclass of metaphorization and metonymization, as the literature reviewed here does not perform this aggregation.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

### 5.3 Change in relation

Metaphorization or metonymization occurs when a word takes on a new meaning that inherits qualities from its original meaning through a figurative relationship the speaker aims to convey (Lakoff and Johnson, 2008, p 3–6). The original meaning is invoked by establishing a material or abstract relation with the initial sense, where the listener is able to resolve the non-literal target concept (Fass, 1988). Material relations are built on physical-world contiguity (e.g., replacing an institution with its location, as in 'Washington decided. . .'). In contrast, abstract relations are formed through conceptual similarity (e.g., applying human body parts to objects, as in 'the arm of the chair').

The theory defining metaphor and metonymy is often informal, relying on subjective perception or overlapping definitions. For example, Choi *et al.* (2021) define metaphor as a case where the 'literal meaning of a word is different from its contextual meaning', which also applies to metonymy. As discussed by Koch (2016, p 42–51), there is no precise rule to define a metonymy. Some computational work, such as Schlechtweg *et al.* (2017), measures a word's entropy to capture metaphorization, but it is not clear that this measure can differentiate a metaphor from a metonymy or a completely new, unrelated sense (e.g., 'Apple' in a technology context).

Given these considerations, we define that a word $w$ has changed in relation if the aggregated sum of the relational similarity ($l$:$S \times S \rightarrow \mathbf{R}$) between its senses changes between corpora. The function $l$ takes two senses, $s_i$ and $s_j$, and returns a real value corresponding to their material or abstract relation (a greater value means a stronger relation). The total relation for a word is:

$$\mathscr{R}(w,t) = \begin{cases} \displaystyle\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} l(s_i, s_j) & s_i, s_j \in \mathscr{S}(w,t) \\ 0 & N \leq 1, \end{cases} \tag{6}$$

where $N = |\mathscr{S}(w,t)|$.

A word $w$ increases in relation between corpora $t_1$ and $t_2$ if $\mathscr{R}(w,t_1) < \mathscr{R}(w,t_2)$, and decreases if the opposite occurs. Since we lack methods to perfectly differentiate material and abstract relations, and since a word can undergo both metaphorization and metonymization, we simplify the interpretation: if a word's figurative usage increases, it 'increases in relation' ; if its figurative usage decreases, it 'decreases in relation'.

### 5.4 Change in orientation

A word's change in orientation (positive or negative connotation) reflects amelioration or pejoration in its usage from one corpus to another. This is measured by analyzing how the collective perception of a word's meaning evolves (Campbell, 2013, p 227–229). A classic example is 'terrific,' which originally meant 'terrible' but shifted to mean 'formidable' or 'impressive.'

While some frameworks like Buechel *et al.* (2016) use a multi-dimensional Valence-Arousal-Dominance setting, most sentiment analysis research focuses on a simpler polarity setting (Cui *et al.* 2023). Therefore, we restrict our formalism to this more general positive/negative view to align with current research and applications (Susanto *et al.* 2020).

We define that a word $w$ had changed in orientation, if for some notion of orientation $\mathfrak{f}$ (feeling), there is a mapping from a word $w$ in a corpora $t$ to an orientation value, for example, in case of positive, neutral, or negative feeling,

$$\mathfrak{f} : V \times T \rightarrow \{-1, 0, +1\}, \tag{7}$$

the orientation changes between corpora $t_1, t_2$,

$$\mathfrak{f}(w,t_1) \neq \mathfrak{f}(w,t_2). \tag{8}$$

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

Analogous to the previous definition, if $\mathfrak{f}(w, t_1) < \mathfrak{f}(w, t_2)$ it suffered an amelioration, and a pejoration if $\mathfrak{f}(w, t_1) > \mathfrak{f}(w, t_2)$.

This task relies on a subjective notion of orientation captured by the function $\mathfrak{f}$. Generally, studies define 'orientation' as the proximity to positive or negative seed words. However, the problem is broader than just polarity, as words can be classified along other emotional dimensions.

### 5.5 Semantic change identification

Core to computational approaches, words in a corpus are not compared directly to one another; instead, they are first represented using one of the many methods reviewed in this survey. A common step taken by all authors is to represent the word 'a' in corpus 'x', denoted as $\mathfrak{R}(a, x)$, and then compare this representation to $\mathfrak{R}(b, y)$. For example, Moss (2020) represents the word 'a' as a Gaussian embedding obtained after training a language model on corpus 'x'. Similarly, Ehmüller *et al.* (2020) represent 'a' using a co-occurrence graph derived from corpus 'x'. Meanwhile, Inoue *et al.* (2022) partition the original corpus into 20-year intervals and, for each resulting corpus 'x', obtain a distribution of topics for the word 'a'.

In this context, we now define in computational terms the task of semantic change characterization. Consider a function $f_{\mathscr{C}}$ that takes two representations $\mathfrak{R}$ (e.g., embeddings, graphs, etc.) of the word $w$ in corpora $t_1, t_2$. We define the problem of semantic change identification as assigning a value ($y$) to this word,

$$f_{\mathscr{C}}(\mathfrak{R}(w, t_1), \mathfrak{R}(w, t_2)) \to y. \tag{9}$$

Schlechtweg and im Walde (2020) proposed two notions of semantic change: binary classification, where $y \in \{0, 1\}$, and graded (continuous) semantic change, where $y \in \mathbf{R}$ is a distance between word representations.[c]

The **binary** classification task aligns with classic LSC literature (Blank, 2012, p 113) and human perception, which is often influenced by sensory thresholds (Smith, 2008, p 34). This view is interested in the gain or loss of senses. There is no general threshold criterion for the binary setting; it's not enough for representations to be different, the magnitude of this difference should align with human judgment.

The **graded** task originates from computational linguistics, where a natural way to compare objects is by measuring the distance between their representations. This approach allows for a more in-depth analysis of change, as the degree of change can be compared across a wide range of contexts.

### 5.6 Semantic change characterization

Deriving from the reviewed literature, we define semantic change characterization as a classification problem where we need to assign identified semantic changes to categories, such as $f_{\mathscr{D}}, f_{\mathscr{R}}, f_{\mathscr{O}}$,

$$f_x(\mathfrak{R}(w, t_1), \mathfrak{R}(w, t_2)) \to y, \qquad x \in \{\mathscr{D}, \mathscr{R}, \mathscr{O}\}. \tag{10}$$

While the first equation only measured some notion of distance between the representations, for characterization, we first apply the type function on the representation, to then compare. For example, in Buechel *et al.* (2016), we have an orientation characterization, for example,

$$f(\mathscr{O}(\mathfrak{R}(w, t_1)), \mathscr{O}(\mathfrak{R}(w, t_2))) \to y, \tag{11}$$

where the $\mathscr{O}$ function is the Valence-Arousal-Dominance framework and they measure the distance $f$ through the $\beta$-coefficient of a linear regression.

[c]While Schlechtweg and im Walde (2020) considered graded change for a particular purpose (i.e., Jensen-Shannon distance), we generalize $y$ to any applicable distance measure.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

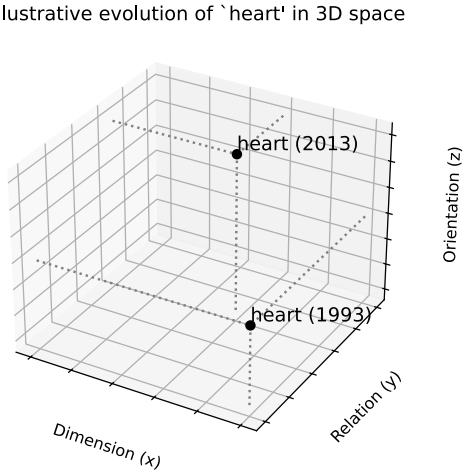


**Figure 11.** Illustrative example of the word 'heart' changing over time. Metaphorization, amelioration and broadening can occur for the same word, depending on the senses it gained/lost.

Again, we set two possibilities of change, binary and graded. In this setting, the binary form allows us to answer interesting questions in an automatic manner, for example, at what point in time the word 'head' started being used in a figurative manner? Another possibility is to track the precise corpus where words were pejorated, and then manually understand the socio-linguistic process of this phenomenon, saving many hours of historiography work.

### *5.7 Framework evaluation*

The proposed framework enables a spatial understanding of a word's change and direct statistical comparison of corpora. In Figure 11, we illustrate this analysis with a possible evolution of the word 'heart' between two corpora from different times (1993 and 2013). A gradual analysis of this evolution can provide a detailed picture across the three poles of change.

To illustrate the framework's practical aspects, we investigate two corpora: SEMCOR (Miller *et al.* 1993) and MASC.[d] SEMCOR is a sense-tagged corpus of 200,000 words created in 1993. MASC is a manually annotated sub-corpus of 500,000 words of modern American English released around 2013.

While the corpora were produced at different times (1993 and 2013), we do not claim this analysis is sufficient to discover all examples of historical LSC. The corpora have narrow coverage and are inadequate for robust diachronic analysis. Additionally, WordNet doesn't cover all existing senses for a given word. We use these datasets for illustrative purposes only. For the purpose of this demonstration, we define corpus $t_1$ as SEMCOR and $t_2$ as MASC, with a time difference of 20 years between the corpora.

SEMCOR and MASC are corpora manually annotated using WordNet senses. We obtain annotations for the poles of change from two lexical resources: ChainNet, a WordNet extension linking senses by metaphor and metonymy, and SentiWordNet, which provides a positive/negative score for every sense in WordNet. This step allows us to skip sense inference and focus on demonstrating the framework. For automatic WordNet annotation, we suggest Navigli *et al.* (2017). For work on annotating data for the poles of change independently of a lexical resource, we suggest de Sá *et al.* (2024). From SEMCOR and MASC, we computed occurrences of 'heart' and 'plane,' two well-known words that have changed semantically.

[d] https://anc.org/data/masc/corpus/

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Table 5.** Distribution of WordNet senses for the word 'heart' in SEMCOR and MASC. The score column indicates a positive score for that sense from SentiWordNet, figurative column indicates if the meaning is figurative or literal

| Sense | Score (+) | Figurative | SEMCOR | MASC |
|---|---|---|---|---|
| heart.n.03 | 0.25 | False | 0.121212 | 0.901408 |
| kernel.n.03 | 0.25 | True | 0.030303 | 0.070423 |
| center.n.01 | 0.00 | True | 0.060606 | 0.000000 |
| heart.n.10 | 0.00 | True | 0.000000 | 0.028169 |
| heart.n.01 | 0.00 | True | 0.439394 | 0.000000 |
| heart.n.02 | 0.00 | False | 0.348485 | 0.000000 |

According to SentiWordNet, 'heart' has two positive senses: courage (heart.n.03) and the vital part of an idea (kernel.n.03). From ChainNet, the absolute literal meanings are the organ (heart.n.02) and courage (heart.n.03). We can observe that in MASC, the literal meaning 'organ' (heart.n.02) has no occurrences, while figurative usage becomes dominant. The word also gained a new sense (heart.n.10, the playing card suit).

From Table 5, we can compute the equations to verify the semantic change across the corpus. We first start with the dimension test. From Equation (5) we have:

$$|\mathscr{S}(w, t_1)| \stackrel{?}{=} |\mathscr{S}(w, t_2)| \tag{12}$$

$$|\mathscr{S}(\text{heart}, \text{SEMCOR})| \stackrel{?}{=} |\mathscr{S}(\text{heart}, \text{MASC})| \tag{13}$$

$$\begin{gathered} |\{\texttt{heart.n.03}, \texttt{kernel.n.03}, \texttt{center.n.01}, \texttt{heart.n.01}, \texttt{heart.n.02}\}| \\ \stackrel{?}{=} \\ |\{\texttt{heart.n.03}, \texttt{kernel.n.03}, \texttt{heart.n.10}\}| \end{gathered} \tag{14}$$

$5 > 3$, we conclude that it narrowed.

To evaluate changes in Orientation, we define the feeling function $\mathfrak{f}$ as the weighted sum of the positive scores of a word's senses. Specifically, we compute $\mathfrak{f}$ as the sum of $\text{positive}(s_i) \cdot p(s_i)$, where $\text{positive}(s_i)$ is the score of sense $s_i$, and $p(s_i)$ is its associated probability. A sense is considered positive if its score is greater than 0.

We use the weighted sum as a proxy for the prototypical sense, assuming that more frequently used senses have a greater influence on the perceived feeling of the word.

$$\mathfrak{f}(w, t) = \frac{1}{N} \sum_{i=1}^{N} p(s_i)\text{positive}(s_i), \quad s_i \in \mathscr{S}(w, t) \tag{15}$$

Plugging the values into our equations, we have:

$$\mathfrak{f}(\text{heart}, \text{SEMCOR}) \stackrel{?}{=} \mathfrak{f}(\text{heart}, \text{MASC}) \tag{16}$$

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Table 6.** Distribution of WordNet senses for the word 'plane' in SEMCOR and MASC. The score column indicates a positive score for that sense from SentiWordNet, and figurative column indicates if the meaning is figurative or literal

| Sense | Score (+) | Figurative | SEMCOR | MASC |
|---|---|---|---|---|
| plane.n.03 | 0.000 | True | 0.046512 | 0.000000 |
| airplane.n.01 | 0.000 | False | 0.488372 | 0.909091 |
| flat.s.01 | 0.375 | False | 0.046512 | 0.000000 |
| plane.v.01 | 0.000 | False | 0.046512 | 0.000000 |
| plane.n.02 | 0.000 | False | 0.372093 | 0.090909 |

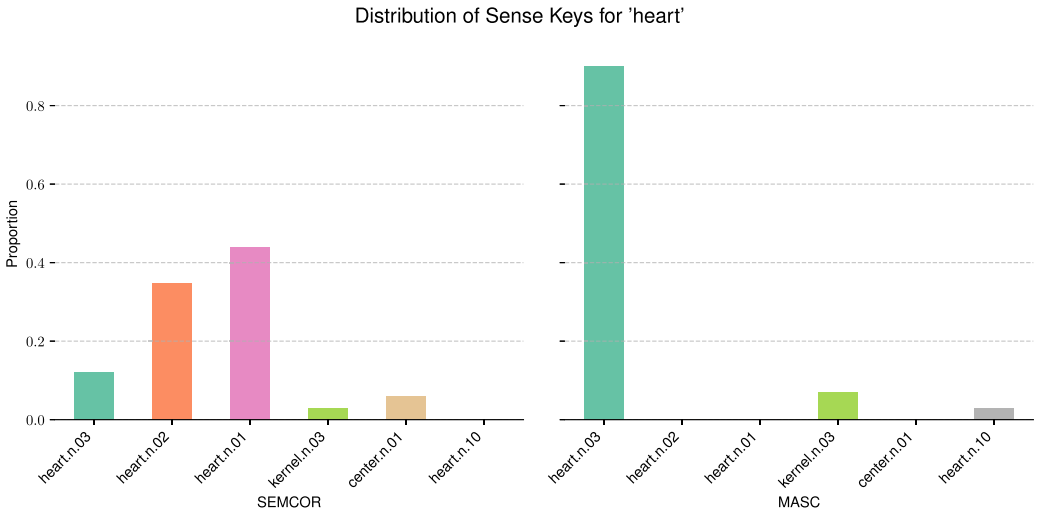


**Figure 12.** Change in meaning distribution for the word 'heart' in SEMCOR and MASC.

$$\begin{aligned} &1 \times 0.12 + 1 \times 0.03 + 0 \times 0.06 + 0 \times 0.0 + 0 \times .44 + 0 \times 0.35 \\ &\stackrel{?}{=} \\ &1 \times 0.9 + 1 \times .07 + 0 \times 0.0 + 0 \times 0.02 + 0 \times 0.0 + 0 \times 0.0 \end{aligned} \tag{17}$$

$0.15 < 0.97$, we observe an amelioration.

Figure 12 illustrates the change in meaning distribution for 'heart.' Under our framework, the word becomes more metaphorical (increasing in Relation, $\mathscr{R}$), narrowed for these corpora (decreasing its Dimension, $\mathscr{D}$), and more positive (increasing in Orientation, $\mathscr{O}$).

For the word 'plane' (Table 6), we note that while SEMCOR reports five distinct senses, MASC presents only two. In MASC, 'plane' has lost its usage as a surface (flat.s.01), a level of existence (plane.n.03), and a type of cut (plane.v.01), while the prototypical usage of 'airplane' has become more prevalent.

For the Relation dimension, we compute $l$ by counting the number of linked senses as described in Equation (6).

$$\mathscr{R}(\text{plane, SEMCOR}) \stackrel{?}{=} \mathscr{R}(\text{plane, MASC}) \tag{18}$$

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

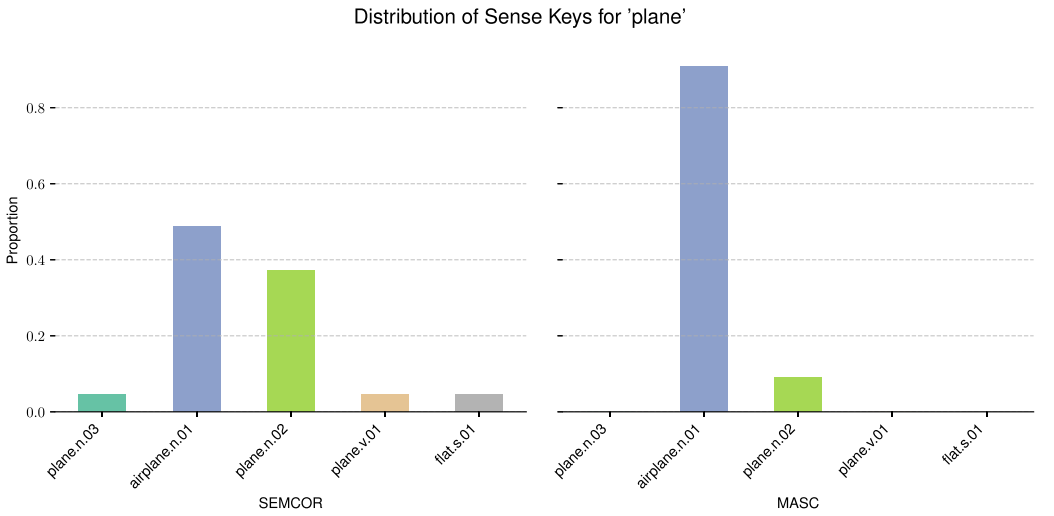


**Figure 13.** Change in meaning distribution for the word 'plane' in SEMCOR and MASC.

$$\begin{aligned} \mathscr{R}(\text{plane, SEMCOR}) = l(\text{plane.n.03, plane.n.02}) = 1 \\ \stackrel{?}{=} \\ \mathscr{R}(\text{plane, MASC}) = 0 \end{aligned} \tag{19}$$

$1 > 0$, so the sense is less metaphorical and decreased in relation.

In Figure 13, we can see the narrowing effect this word suffered in this corpus. The decreased usage as a positive meaning (flat.s.01) causes a decrease in its Orientation ($\mathscr{O}$).

## 6. Discussion

In this section, we synthesize the reviewed approaches for change characterization and discuss their limitations according to the dimension, relation, and orientation aspects we have introduced and formalized.

### *6.1 Dimension*

The majority of studies found in this survey propose methods to characterize how words change in dimension, while the other poles are less studied. Change in dimension seems easier to detect because it does not rely on subjective interpretation, such as relation or feeling, but mainly on a countable quantity (see Section 5). Work along this axis often relies on word sense induction, a research area with well-developed tooling and many data sources, such as dictionaries and synsets, used for supervised learning.

The key factor differentiating the approaches is the word representation method. We identified four main types – frequency, topics, graphs, and embeddings – each with different trade-offs in implementation cost, representation power, and interpretability. It is difficult to compare the effectiveness of each method for the current state of the art, given the qualitative nature of reported results, the use of different corpora, and the need for external knowledge. However, our analysis reveals that frequency- and embedding-based methods appear to produce better results for identifying neologisms (new words), while topic- and graph-based approaches are more effective at detecting sememes (new senses) over time.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

A limitation for approaches based on dictionaries is the temporal delay between a neologism's first appearance and its inclusion in dictionaries. Another limitation is the presence of metaphorical senses at the same level as literal ones, increasing the complexity of distinguishing between 'dimension' and 'relation.' Some authors prefer to group these two poles into one (Tang *et al.* 2013). For topic- and graph-based approaches, the limitation is that interpreting the meaning of a topic or a graph branch is not always clear and can require expert intervention.

### 6.2 Relation

Outside of this LSC survey, a vast body of work proposes methods to predict if a word in a sentence has a metaphorical sense (Heintz *et al.* 2013; Shutova and Sun 2013; Hovy *et al.* 2013). These approaches analyze texts but do not consider temporal information, so they cannot determine if a word has undergone metaphorization over time. Some works claim to detect the metaphorization of words (e.g., Tredici, Nissim, and Zaninello 2016), but they rely on the general hypothesis that any sense change implies a metaphorical change, which still needs to be demonstrated.

For metonymization, we could not find studies that performed metonymy identification in a comparative setting. However, related recent studies investigate metonymy usage with various methods for material relation identification (Nastase *et al.* 2012; Soares *et al.* 2019; Mathews and Strube 2021). We expect that future investigations can extend these approaches to detect word metonymization.

Among the four meaning representation methods, frequency-based approaches were the first investigated for relation characterization. However, their underlying hypothesis doesn't differentiate between new metaphorical/metonymical usage and new homonymous usage, making interpretation difficult. More recent works employed embeddings, but interpreting vectors in high-dimensional space remains a challenge. Finally, graphs, well-known for representing relations, remain under-investigated for identifying changes in relations.

Tang *et al.* (2013) mainly rely on dictionaries of heuristics to infer metaphoricity, which served for qualitative analysis, but still require more extensive validation. A single work used ground-truth annotations for metaphors (Schlechtweg *et al.* 2017), but did not report a final precision of the proposed methodology.

As noted, studies on the 'relation pole' are few and inconclusive. The research community's limited interest in this topic could be justified by three reasons: (I) the difficulty of formally defining metaphor and metonymy from a computational perspective, (II) the rarity and insufficiency of annotated corpora for relation detection, and (III) the nonexistence of standard metrics to measure and compare approaches.

### 6.3 Orientation

Interest in orientation analysis is increasing with the wide adoption of social networks. Orientation is usually determined using a feeling function (ϱ) that relies on seed words and the assumption that these seed words don't vary over time. As we illustrated, word meanings evolve and can differ between communities. Most works found in this survey addressed the change in orientation for particular words Miller-Lewis *et al.* (2021); Xiao *et al.* (2023).

The example of 'sick' shows that a word can have two opposite orientations depending on the context. Creating corpora that include all these variations with statistical significance is a complex task. The annotation process can suffer from many biases, both from annotators and data. Selected seed words may not generalize across all contexts or long time spans, and a series of misleading results can occur, as presented in Antoniak and Mimno (2021).

Of the four meaning representation methods, embedding- and frequency-based approaches are the most common. Few works use graphs, and we could not find any approach that uses topic analysis for detecting word orientation. The preference for embeddings and frequency is likely

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

because current approaches compute distances or co-occurrences with seed words, which is easier with these methods.

While there is a huge body of work on sentence-level sentiment analysis, fewer studies infer sentiment at the word level (Guo, Yu, and Wang 2021), and only a small fraction of those do so in a comparative manner, as we have assessed in this survey. While works in topics provide lexical entries that could be used similarly to what was done in embeddings, this was not investigated so far.

An important outcome of this survey is that there is some harmonization in how a change in feeling is defined and applied. However, we did not observe a harmonized approach to error analysis for the feeling function. Some works adopt a binary orientation function (positive, negative), while others use multidimensional orientations (valence, arousal, dominance) (Buechel *et al.* 2016). Further study is needed to analyze which function (ƒ) is most suitable for characterizing change.

A converging point from the many approaches analyzed is the difficulty of evaluating and comparing results because, for most tasks, no gold standard exists. Many papers have raised this problem (Schlechtweg *et al.* 2017), but some initiatives to create reference datasets are underway. Another point is multilingualism; the vast majority of experiments use only English documents, making it difficult to generalize methods to other languages. The recent emergence of new-generation language models has opened up new possibilities for identifying and characterizing semantic changes, but this area of research is still underexplored.

We begin this survey with five objectives, which we believe have been achieved:

1. We discussed the limitations of current representations and methodologies in this section.
2. We evaluated the strengths of various approaches and the complexity of their computational interpretation in Table 4.
3. We proposed formal definitions in Section 5 to reduce ambiguity.
4. We demonstrate our proposed formalization with real data in Subsection 5.7.

## 7. Conclusion

In this survey, we reviewed current work on semantic change characterization with respect to the mainstream typology (broadening/narrowing, amelioration/pejoration, and metaphorization/metonymization). We observed four major approaches to representing meaning – frequency, topics, embeddings, and graphs – and their respective strengths and drawbacks.

After analyzing the literature, we observed that many tasks are underexplored or addressed without systematic evaluation. Therefore, we introduced a formal definition for the task of semantic change characterization, a more complex instance of the semantic change identification problem. This new task is extremely relevant for the field of LSC, as this study and other surveys have argued.

The suggested tasks are crucial for both linguistic change research and NLP. Numerous applications, such as sentiment analysis (orientation), semantic similarity (relation), and word sense inference (dimension), call for a deeper understanding of these representational characteristics. Additionally, awareness of how these changes affect word representations leads to better knowledge and interpretability of these systems. For word embeddings specifically, a better understanding of how these changes affect vector space geometries could improve our understanding of the semantics of vectors.

As future work, we expect to investigate the relation pole to better differentiate categories of change and further explore how the four representational methods relate to the three poles. Another important goal is the consolidation of datasets and metrics by which semantic change characterization can be compared across methodologies. In this survey, we performed an initial

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

investigation of existing sense-annotated corpora; in future studies, we expect to analyze words in more comprehensive corpora.

Finally, this survey has limitations in its coverage of LSC and computational linguistics. Since the study of word meaning involves multiple, rapidly evolving fields, we focused on key works related to computational characterization and traditional meaning representations. The rise of new-generation language models and new theories of language development offers further opportunities for studying semantic change, but this area is still not fully explored.

**Acknowledgements.** The research reported in this publication was supported by the Luxembourg National Research Fund (FNR), project D4H grant number PRIDE21/16758026.

**Use of AI tools.** The production of this manuscript used AI tools solely to improve grammar and readability, as English is not the authors' native language.

## References

**Allan K.** (2013). *The Oxford Handbook of the History of Linguistics*. OUP Oxford.

**Allan K.** and **Robinson J. A.** (2012). *Current Methods in Historical Semantics*. Berlin: De Gruyter Mouton.

**Antoniak M.** and **Mimno D. M.** (2021). Bad seeds: Evaluating lexical methods for bias measurement. In *Annual Meeting of the Association for Computational Linguistics*.

**Ballout M.**, **Dedert A.**, **Abdelmoneim N.**, **Krumnack U.**, **Heidemann G.** and **Kühnberger K.-U.** (2024). Fool me if you can! an adversarial dataset to investigate the robustness of LMs in word sense disambiguation. In *Conference on Empirical Methods in Natural Language Processing*.

**Barnes J.**, **Kurtz R.**, **Oepen S.**, **Ovrelid L.** and **Velldal E.** (2021). Structured sentiment analysis as dependency graph parsing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, pp. 3387–3402.

**Biemann C.** (2006). Chinese whispers – an efficient graph clustering algorithm and its application to natural language processing problems. In *Proceedings of TextGraphs: The First Workshop on Graph Based Methods for Natural Language Processing*. New York City: Association for Computational Linguistics, pp. 73–80.

**Blank A.** (2012). *Prinzipien Des Lexikalischen Bedeutungswandels Am Beispiel der Romanischen Sprachen*, vol. **285**. Walter de Gruyter.

**Blank A.** (2003). Polysemy in the lexicon and in discourse. In *Polysemy: Flexible Patterns of Meaning in Mind and Language* .

**BNC** (2025). BNC. Available at http://www.natcorp.ox.ac.uk/ (Accessed 29/07/2025).

**Bochkarev V. V.**, **Khristoforov S.**, **Shevlyakova A. V.** and **Solovyev V. D.** (2022). Neural network algorithm for detection of new word meanings denoting named entities. *IEEE Access* **10**, 68499–68512.

**Buechel S.**, **Hellrich J.** and **Hahn U.** (2016). Feelings from the past – adapting affective lexicons for historical emotion analysis. In *LT4DH@COLING*.

**Camacho-Collados J.** and **Pilehvar M. T.** (2018). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research* **63**, 743–788.

**Campbell L.** (2013). *Historical Linguistics*. Edinburgh University Press.

**Carpuat M.**, **Daumé H.**, **Henry K.**, **Irvine A.**, **Jagarlamudi J.** and **Rudinger R.** (2013). Sensespotting: Never let your parallel data tie you to an old domain. In *Annual Meeting of the Association for Computational Linguistics*.

**Cassotti P.**, **De Pascale S.** and **Tahmasebi N.** (2024). Using synchronic definitions and semantic relations to classify semantic change types. In Ku, L.-W., Martins, A., and Srikumar, V. (eds), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Bangkok, Thailand: Association for Computational Linguistics, pp. 4539–4553.

**Cassotti P.**, **Siciliani L.**, **Degemmis M.**, **Semeraro G.** and **Basile P.** (2023). XL-LEXEME: WiC Pretrained Model for Cross-Lingual Lexical Semantic Change. In *Annual Meeting of the Association for Computational Linguistics*.

**Choi M.**, **Lee S.**, **Choi E.**, **Park H.**, **Lee J.**, **Lee D.** and **Lee J.** (2021). MelBERT: Metaphor detection via contextualized late interaction using metaphorical identification theories. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, pp. 1763–1773. Online.

**CLMET** (2025). CLMET. Available at https://perswww.kuleuven.be/~u0044428/clmet3_0.htm (Accessed 29/07/2025).

**COHA** (2025). COHA. Available at https://www.english-corpora.org/coha/ (Accessed 29/07/2025).

**Conneau A.**, **Lample G.**, **Ranzato M.**, **Denoyer L.** and **J'egou H.** (2018). Word translation without parallel data. In *International Conference on Learning Representations*.

**Cook P.**, **Lau J. H.**, **Rundell M.**, **McCarthy D.** and **Baldwin T.** (2013). A lexicographic appraisal of an automatic approach for detecting new word-senses. In *Proceedings of eLex*.

**Cook P.** and **Stevenson S.** (2010). Automatically identifying changes in the semantic orientation of words. In *International Conference on Language Resources and Evaluation*.

**Cui J.**, **Wang Z.**, **Ho S.-B.** and **Cambria E.** (2023). Survey on sentiment analysis: Evolution of research methods and topics. *Artificial Intelligence Review* **56**, 1–42.

**Danilevsky M.**, **Qian K.**, **Aharonov R.**, **Katsis Y.**, **Kawas B.** and **Sen P.** (2020). A survey of the state of explainable ai for natural language processing. In *AACL*.

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

**de Sá J. M. C.**, **Silveira M. D.** and **Pruski C.** (2024). Semantic change characterization with llms using rhetorics. In *The Proceedings for the 6th International Workshop on Computational Approaches to Language Change (LChange'26)*. Association for Computational Linguistics, pp. 110–123.

**Devlin J.**, **Chang M.-W.**, **Lee K.** and **Toutanova K.** (2019). Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186.

**DTA** (2025). DTA. Available at https://www.deutschestextarchiv.de/ (Accessed 29/07/2025).

**Dubossarsky H.**, **Weinshall D.** and **Grossman E.** (2017). Outta control: Laws of semantic change and inherent biases in word representation models. In *Conference on Empirical Methods in Natural Language Processing*.

**Ehmüller J.**, **Kohlmeyer L.**, **McKee H.**, **Paeschke D.**, **Repke T.**, **Krestel R.** and **Naumann F.** (2020). Sense tree: Discovery of new word senses with graph-based scoring. In *Lernen, Wissen, Daten, Analysen*.

**Emms M.** and **Jayapal A. K.** (2016). Dynamic generative model for diachronic sense emergence detection. In *International Conference on Computational Linguistics*.

**Fass D.** (1988). Metonymy and metaphor: What's the difference? In International Conference on Computational Linguistics.

**Fonteyn L.** and **Manjavacas E.** (2021). Adjusting scope: A computational approach to case-driven research on semantic change. In *Workshop on Computational Humanities Research*.

**Frermann L.** and **Lapata M.** (2016). A bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics* **4**, 31–45.

**Gao T.**, **Yao X.** and **Chen D.** (2021). SimCSE: Simple Contrastive Learning of Sentence Embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 6894–6910.

**Geeraerts D.** (1997). *Diachronic Prototype Semantics: A Contribution to Historical Lexicology*. Oxford University Press.

**Giulianelli M.**, **Luden I.**, **Fernández R.** and **Kutuzov A.** (2023). Interpretable word sense representations via definition generation: The case of semantic change analysis. In *Annual Meeting of the Association for Computational Linguistics*.

**Giulianelli M.**, **Tredici M. D.** and **Fern'andez R.** (2020). Analysing lexical semantic change with contextualised word representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3960–3973.

**Gladkova A.** and **Drozd A.** (2016). Intrinsic evaluations of word embeddings: What can we do better? In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP, RepEval@ACL 2016, Berlin, Germany, August 2016*. Association for Computational Linguistics, pp. 36–42.

**Google N-Gram Book** (2025). Google N-Gram Book corpus. Available at https://commondatastorage.googleapis.com/books/syntactic-ngrams/index.html (Accessed 29/07/2025).

**Günther F.**, **Rinaldi L.** and **Marelli M.** (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science* **14**, 1006–1033.

**Guo X.**, **Yu W.** and **Wang X.** (2021). An overview on fine-grained text sentiment analysis: Survey and challenges. *Journal of Physics: Conference Series* **1757**, 012038.

**Hamilton W. L.**, **Clark K.**, **Leskovec J.** and **Jurafsky D.** (2016a). Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Conference on Empirical Methods in Natural Language Processing, pp. 595–605.

**Hamilton W. L.**, **Leskovec J.** and **Jurafsky D.** (2016b). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1489–1501.

**Hammarström H.**, **Rönchen P.**, **Elgh E.** and **Wiklund T.** (2019). On computational historical linguistics in the 21st century. *Theoretical Linguistics* **45**, 233–245.

**He P.**, **Gao J.** and **Chen W.** (2021). Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *International Conference in Learning Representations*.

**Heintz I.**, **Gabbard R.**, **Srivastava M.**, **Barner D.**, **Black D.**, **Friedman M.** and **Weischedel R. M.** (2013). Automatic extraction of linguistic metaphors with LDA topic modeling. In *Proceedings of the First Workshop on Metaphor in NLP*, pp. 58–66.

**Hellrich J.**, **Buechel S.** and **Hahn U.** (2018). Modeling word emotion in historical language: Quantity beats supposed stability in seed word selection. In *LaTeCH@NAACL-HLT*.

**Hellrich J.** and **Hahn U.** (2016). Bad company – neighborhoods in neural embedding spaces considered harmful. In **Calzolari, N., Matsumoto, Y., and Prasad, R.**, editors, *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11–16, 2016, Osaka, Japan*, ACL, pp. 2785–2796.

**Helsinki Corpus** (2025). Helsinki Corpus. Available at https://varieng.helsinki.fi/CoRD/corpora/HelsinkiCorpus/ (Accessed 29/07/2025).

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Hengchen S.**, **Tahmasebi N.**, **Schlechtweg D.** and **Dubossarsky H.** (2021). *Challenges for Computational Lexical Semantic Change*. Language Science Press.

**Hock H. H.** and **Joseph B. D.** (2019). *Chapter 7: Semantic Change*. Berlin, Boston: De Gruyter Mouton.

**Hovy D.**, **Shrivastava S.**, **Jauhar S. K.**, **Sachan M.**, **Goyal K.**, **Li H.**, **Sanders W. E.** and **Hovy E. H.** (2013). Identifying metaphorical word use with tree kernels. In *Proceedings of the First Workshop on Metaphor in NLP*.

**Hu R.**, **Li S.** and **Liang S.** (2019). Diachronic sense modeling with deep contextualized word embeddings: An ecological view. In *Annual Meeting of the Association for Computational Linguistics*.

**Inoue S.**, **Komachi M.**, **Ogiso T.**, **Takamura H.** and **Mochihashi D.** (2022). Infinite scan: An infinite model of diachronic semantic change. In *Conference on Empirical Methods in Natural Language Processing*.

**Ito T.**, **Tsubouchi K.**, **Sakaji H.**, **Yamashita T.** and **Izumi K.** (2020). Word-level contextual sentiment analysis with interpretability. In *AAAI Conference on Artificial Intelligence*.

**Jatowt A.** and **Duh K.** (2014). A framework for analyzing semantic change of words across time. In *IEEE/ACM Joint Conference on Digital Libraries*, pp. 229–238.

**Koch P.** (2016). Meaning change and semantic shifts. In *The Lexical Typology of Semantic Shifts* **58**, 21–66.

**Kubhist2** (2025). Kubhist2. Available at https://spraakbanken.gu.se/en/resources/kubhist2 (Accessed 29/07/2025).

**Kutuzov A.**, **Øvrelid L.**, **Szymanski T.** and **Velldal E.** (2018). Diachronic word embeddings and semantic shifts: A survey. In *International Conference on Computational Linguistics*.

**Kutuzov A.**, **Velldal E.** and **Øvrelid L.** (2022). Contextualized embeddings for semantic change detection: Lessons learned. *Northern European Journal of Language Technology*, 8.

**Lakoff G.** and **Johnson M.** (2008). *Metaphors We Live by*. University of Chicago press.

**Lampeter Corpus** (2025). Lampeter Corpus. Available at https://varieng.helsinki.fi/CoRD/corpora/LC/ (Accessed 29/07/2025).

**Lau J. H.**, **Cook P.**, **McCarthy D.**, **Gella S.** and **Baldwin T.** (2014). Learning word sense distributions, detecting unattested senses and identifying novel senses using topic models. In *Annual Meeting of the Association for Computational Linguistics*.

**Lau J. H.**, **Cook P.**, **McCarthy D.**, **Newman D.** and **Baldwin T.** (2012). Word sense induction for novel sense detection. In *Conference of the European Chapter of the Association for Computational Linguistics*.

**Lautenschlager J.**, **Sköldberg E.**, **Hengchen S.** and **Schlechtweg D.** (2024). Detection of non-recorded word senses in English and Swedish. arXiv preprint. ArXiv, abs/2403.02285.

**Lin D.** (1998). An information-theoretic definition of similarity. In *International Conference on Machine Learning*.

**Liu H.**, **Tam D.**, **Muqeeth M.**, **Mohta J.**, **Huang T.**, **Bansal M.** and **Raffel C.** (2022). Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Advances in Neural Information Processing Systems* **35**, 1950–1965.

**Mathew B.**, **Maity S. K.**, **Sarkar P.**, **Mukherjee A.** and **Goyal P.** (2017). Adapting predominant and novel sense discovery algorithms for identifying corpus-specific sense differences. In *Proceedings of TextGraphs-11: The Workshop on Graph-based Methods for Natural Language Processing*, Vancouver, Canada: Association for Computational Linguistics, pp. 11–20.

**Mathews K. A.** and **Strube M.** (2021). Impact of target word and context on end-to-end metonymy detection. arXiv preprint. ArXiv, abs/2112.03256.

**Maudslay R. H.** and **Teufel S.** (2022). Metaphorical polysemy detection: Conventional metaphor meets word sense disambiguation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pp. 65–77.

**McCarthy D.**, **Koeling R.**, **Weeds J.** and **Carroll J.** (2004). Finding predominant word senses in untagged text. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, Barcelona, Spain, pp. 279–286.

**Mikolov T.**, **Chen K.**, **Corrado G. S.** and **Dean J.** (2013). Efficient estimation of word representations in vector space. In *International Conference on Learning Representations*.

**Miller G. A.**, **Leacock C.**, **Tengi R.** and **Bunker R.** (1993). A semantic concordance. In Human Language Technology - The Baltic Perspectiv.

**Miller-Lewis L. R.**, **Lewis T. W.**, **Tieman J. J.**, **Rawlings D.**, **Parker D.** and **Sanderson C. R.** (2021). Words describing feelings about death: A comparison of sentiment for self and others and changes over time. *PloS one* **16**, e0242848.

**Min S.**, **Lyu X.**, **Holtzman A.**, **Artetxe M.**, **Lewis M.**, **Hajishirzi H.** and **Zettlemoyer L.** (2022). Rethinking the role of demonstrations: What makes in-context learning work? In *Conference on Empirical Methods in Natural Language Processing*.

**Mitchell J.** and **Lapata M.** (2010). Composition in distributional models of semantics. *Cognitive Science* **34**, 1388–1429.

**Mitra S.**, **Mitra R.**, **Maity S. K.**, **Riedl M.**, **Biemann C.**, **Goyal P.** and **Mukherjee A.** (2015). An automatic approach to identify word sense changes in text media across timescales. *Natural Language Engineering* **21**, 773–798.

**Mohammed E. T.** (2009). Polysemy as a lexical problem in translation. *Online Submission* **55**, 1–19.

**Montanelli S.** and **Periti F.** (2023). A survey on contextualised semantic shift detection. arXiv preprint. ArXiv, abs/2304.01666.

**Moss A.** (2020). Detecting lexical semantic change using probabilistic gaussian word embeddings. [Master's thesis, Uppsala University] Retrieved from https://www.essays.se/essay/be0f7d4e4a/.

**Nastase V.**, **Judea A.**, **Markert K.** and **Strube M.** (2012). Local and global context for supervised and unsupervised metonymy resolution. In *Conference on Empirical Methods in Natural Language Processing*.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Navigli R.**, **Camacho-Collados J.** and **Raganato A.** (2017). Word sense disambiguation: A unified evaluation framework and empirical comparison. In *Conference of the European Chapter of the Association for Computational Linguistics*.

**Neubig G.** and **Hu J.** (2018). Rapid adaptation of neural machine translation to new languages. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 875–880.

**Nevalainen T.** and **Traugott E. C.** (2016). *The Oxford Handbook of the History of English*. Oxford University Press.

**Nivre J.**, **de Marneffe M.-C.**, **Ginter F.**, **Hajivc J.**, **Manning C. D.**, **Pyysalo S.**, **Schuster S.**, **Tyers F. M.** and **Zeman D.** (2020). Universal dependencies v2: An evergrowing multilingual treebank collection. In *International Conference on Language Resources and Evaluation*.

**NYT Archives** (2025). NYT archives. Available at https://archive.nytimes.com/www.nytimes.com/ref/membercenter/nytarchive.html (Accessed 29/07/2025).

**Pedinotti P.** and **Lenci A.** (2020). Don't invite BERT to drink a bottle: Modeling the interpretation of metonymies using BERT and distributional representations. In *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain: International Committee on Computational Linguistics, pp. 6831–6837.

**People's Daily Newspaper** (2025). People's Daily Newspaper. Available at https://www.oriprobe.com/peoplesdaily.shtml (Accessed 29/07/2025).

**Periti F.**, **Cassotti P.**, **Dubossarsky H.** and **Tahmasebi N.** (2024). Analyzing semantic change through lexical replacements. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pp. 4495–4510.

**Pinker S.** (2003). *The Language Instinct: How the Mind Creates Language*. Penguin UK.

**Pinker S.** and **Bloom P. J.** (1990). Natural language and natural selection. *Behavioral and Brain Sciences* **13**, 725–726.

**Pivovarova L.** and **Kutuzov A.** (2021). Rushifteval: a shared task on semantic shift detection for Russian. arXiv preprint. ArXiv, abs/2106.08294.

**Polleres A.**, **Pernisch R.**, **Bonifati A.**, **Dell'Aglio D.**, **Dobriy D.**, **Dumbrava S.**, **Etcheverry L.**, **Ferranti N.**, **Hose K.**, **Jiménez-Ruiz E.**, **Lissandrini M.**, **Scherp A.**, **Tommasini R.** and **Wachs J.** (2023). How does knowledge evolve in open knowledge graphs? *Transactions on Graph Data and Knowledge* **1**, 11:1–11:59.

**Project Gutenberg** (2025). Project Gutenberg. Available at https://www.gutenberg.org/ (Accessed 29/07/2025).

**Raffel C.**, **Shazeer N. M.**, **Roberts A.**, **Lee K.**, **Narang S.**, **Matena M.**, **Zhou Y.**, **Li W.** and **Liu P. J.** (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research* **21**, 140:1–140:67.

**Reagan A. J.**, **Danforth C. M.**, **Tivnan B. F.**, **Williams J. R.** and **Dodds P. S.** (2017). Sentiment analysis methods for understanding large-scale texts: A case for using continuum-scored words and word shift graphs. *EPJ Data Science* **6**, 1–21.

**Reimers N.** and **Gurevych I.** (2019). Sentence-BERT: Sentence embeddings using siamese bert-networks. In *Conference on Empirical Methods in Natural Language Processing*.

**Rilling J. K.**, **Glasser M. F.**, **Jbabdi S.**, **Andersson J.** and **Preuss T. M.** (2012). Continuity, divergence, and the evolution of brain language pathways. *Frontiers in Evolutionary Neuroscience* **3**, 11.

**Rosenfeld A.** and **Erk K.** (2018). Deep neural models of semantic shift. In *North American Chapter of the Association for Computational Linguistics*.

**Ryskina M.**, **Rabinovich E.**, **Berg-Kirkpatrick T.**, **Mortensen D. R.** and **Tsvetkov Y.** (2020). Where new words are born: Distributional semantic analysis of neologisms and their semantic neighborhoods. In *Proceedings of the Society for Computation in Linguistics*, pp. 367–376.

**Sagi E.**, **Kaufmann S.** and **Clark B.** (2009). Semantic density analysis: Comparing word meaning across time and phonetic space. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics, GEMS '09*. Association for Computational Linguistics, vol. **09**, pp. 104–111.

**Schlechtweg D.**, **Eckmann S.**, **Santus E.**, **im Walde S. S.** and **Hole D.** (2017). German in flux: Detecting metaphoric change via word entropy. In *Conference on Computational Natural Language Learning*.

**Schlechtweg D.**, **Hätty A.**, **Tredici M. D.** and **im Walde S. S.** (2019). A wind of change: Detecting and evaluating lexical semantic change across times and domains. In *Annual Meeting of the Association for Computational Linguistics*.

**Schlechtweg D.** and **im Walde S. S.** (2020). Simulating lexical semantic change from sense-annotated data. arXiv preprint. ArXiv, abs/2001.03216.

**Schlechtweg D.**, **McGillivray B.**, **Hengchen S.**, **Dubossarsky H.** and **Tahmasebi N.** (2020). SemEval-2020 task 1: Unsupervised lexical semantic change detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. Barcelona: International Committee for Computational Linguistics, pp. 1–23.

**Sherwood C. C.**, **Subiaul F.** and **Zawidzki T. W.** (2008). A natural history of the human mind: tracing evolutionary changes in brain and cognition. *Journal of Anatomy* **212**, 426–454.

**Shoemark P.**, **Liza F. F.**, **Nguyen D.**, **Hale S. A.** and **McGillivray B.** (2019). Room to Glo: A systematic comparison of semantic change detection approaches with word embeddings. In *Conference on Empirical Methods in Natural Language Processing*.

**Shutova E.** and **Sun L.** (2013). Unsupervised metaphor identification using hierarchical graph factorization clustering. In *North American Chapter of the Association for Computational Linguistics*.

**Smith C. U. M.** (2008). *Biology of Sensory Systems*. John Wiley & Sons.

https://doi.org/10.1017/nlp.2026.10023 Published online by Cambridge University Press

**Soares L. B.**, **FitzGerald N.**, **Ling J.** and **Kwiatkowski T.** (2019). Matching the blanks: Distributional similarity for relation learning. In *Annual Meeting of the Association for Computational Linguistics*.

**Socher R.**, **Perelygin A.**, **Wu J.**, **Chuang J.**, **Manning C. D.**, **Ng A.** and **Potts C.** (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Conference on Empirical Methods in Natural Language Processing*.

**Susanto Y.**, **Livingstone A. G.**, **Ng B. C.** and **Cambria E.** (2020). The hourglass model revisited. *IEEE Intelligent Systems* **35**, 96–102.

**Tahmasebi N.**, **Borin L.** and **Jatowt A.** (2018). *Survey of computational approaches to lexical semantic change*. Language Science Press.

**Tahmasebi N.** and **Risse T.** (2017). Finding individual word sense changes and their delay in appearance. In *Recent Advances in Natural Language Processing*.

**Takamura H.**, **Nagata R.** and **Kawasaki Y.** (2017). Analyzing semantic change in Japanese loanwords. In *Conference of the European Chapter of the Association for Computational Linguistics*.

**Tang X.** (2018). A state-of-the-art of semantic change computation. *Natural Language Engineering* **24**, 649–676.

**Tang X.**, **Qu W.** and **Chen X.** (2013). Semantic change computation: A successive approach. *World Wide Web-Internet and Web Information Systems* **19**, 375–415.

**Tang X.**, **Zhou Y.**, **Aida T.**, **Sen P.** and **Bollegala D.** (2023). Can word sense distribution detect semantic changes of words? In *Findings of the Association for Computational Linguistics: EMNLP*, pp. 3575–3590.

**Tay Y.**, **Dehghani M.**, **Tran V. Q.**, **García X.**, **Wei J.**, **Wang X.**, **Chung H. W.**, **Bahri D.**, **Schuster T.**, **Zheng H. S.**, **Zhou D.**, **Houlsby N.** and **Metzler D.** (2022). UL2: Unifying language learning paradigms. In *International Conference on Learning Representations*.

**Teraoka T.** (2016). Metonymy analysis using associative relations between words. In *International Conference on Language Resources and Evaluation*.

**Traugott E. C.** (2017). Semantic change. *Oxford Research Encyclopedia of Linguistics*.

**Tredici M. D.**, **Nissim M.** and **Zaninello A.** (2016). Tracing metaphors in time through self-distance in vector spaces. arXiv preprint. ArXiv, abs/1611.03279.

**Uban A.-S.**, **Ciobanu A. M.** and **Dinu L. P.** (2021). Cross-lingual laws of semantic change. In *Computational Approaches to Semantic Change*, vol. **6**, 219.

**Vulic I.**, **Ruder S.** and **Søgaard A.** (2020). Are all good word vector spaces isomorphic? In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 3178–3192.

**Wang B.**, **Buccio E. D.** and **Melucci M.** (2021). Word2Fun: Modelling words as functions for diachronic word representation. In *Neural Information Processing Systems*.

**Wang Y.-S.**, **Wu A.** and **Neubig G.** (2022). English contrastive learning can learn universal cross-lingual sentence embeddings. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 9122–9133.

**Wankhade M.**, **Rao A. C. S.** and **Kulkarni C.** (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review* **55**, 5731–5780.

**Warriner A. B.**, **Kuperman V.** and **Brysbaert M.** (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* **45**, 1191–1207.

**Wittgenstein L.** (2009). *Philosophical Investigations*. John Wiley & Sons.

**Wu S.** and **Dredze M.** (2019). Beto, bentz, becas: The surprising cross-lingual effectiveness of bert. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 833–844.

**Xiao Y.**, **Baes N.**, **Vylomova E.** and **Haslam N.** (2023). Have the concepts of 'anxiety' and 'depression' been normalized or pathologized? A corpus study of historical semantic change. *PLOS ONE* **18**, e0288027.

**Xie J. Y.**, **Junior R. F. P.**, **Hirst G.** and **Xu Y.** (2019). Text-based inference of moral sentiment change. In *Conference on Empirical Methods in Natural Language Processing*.

**Yüksel A.**, **Ugurlu B.** and **Koç A.** (2021). Semantic change detection with gaussian word embeddings. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **29**, 3349–3361.

**Zhang X.**, **van de Meent J.-W.** and **Wallace B. C.** (2021). Disentangling representations of text by masking transformers. In *Conference on Empirical Methods in Natural Language Processing*.

**Zhou K.**, **Ethayarajh K.** and **Jurafsky D.** (2021). Frequency-based distortions in contextualized word embeddings. arXiv preprint. ArXiv, abs/2104.08465.

**Zipf G. K.** (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press, Inc.

**Cite this article:** de Sá JMC, Da Silveira M and Pruski C (2026). Survey in characterizing semantic change. *Natural Language Processing* **32**, 233–266. https://doi.org/10.1017/nlp.2026.10023