

# Modeling Changing Scientific Concepts with Complex Networks: A Case Study on the Chemical Revolution

Sofía Aguilar-Valdez and Stefania Degaetano-Ortlieb

Language Science and Technology Department, Saarland University

sofia.aguilar@uni-saarland.de, s.degaetano@mx.uni-saarland.de

## Abstract

While context embeddings produced by LLMs can be used to estimate conceptual change, these representations are often not interpretable nor time-aware. Moreover, bias augmentation in historical data poses a non-trivial risk to researchers in the Digital Humanities. Hence, to model reliable concept trajectories in evolving scholarship, in this work we develop a framework that represents prototypical concepts through complex networks based on topics. Utilizing the Royal Society Corpus, we analyzed two competing theories from the Chemical Revolution (phlogiston vs. oxygen) as a case study to show that onomasiological change is linked to higher entropy and topological density, indicating increased diversity of ideas and connectivity effort.

## 1 Introduction

Language change is not random; it is driven by shifting communicative goals, social structures, and domain-specific conventions (Hamilton et al., 2016b; Gries and Hilpert, 2008; Blank, 2013). While various methods exist to model change in language use, such as lexical semantic change through word meaning representations from LLMs, interpretability remains a difficult task (Periti and Montanelli, 2024). This impedes domain experts' ability to qualitatively assess change and adapt models to their specific needs (Beck, 2024).

We model conceptual change through language use in the context of scientific revolutions, where shifts in language use reflect broader epistemic transitions. Specifically, we want to analyze how the lexemes referring to a concept evolve after discoveries (Zgusta, 2011), which poses the challenges of quantifying concepts with adaptable structures and predicting what conceptual evolution mechanisms are associated with epistemic drifts. To this end, we propose a graph-based approach that models scientific texts as concept networks, organizing

1780

1800

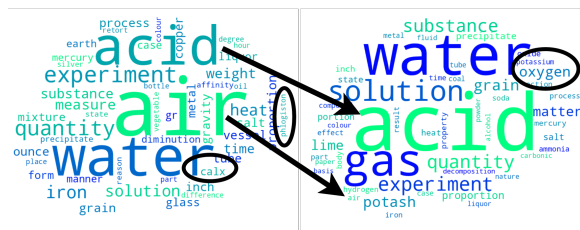


Figure 1: **Diachronic prototypical concepts.** This schematic illustrates movements in central and peripheral readings (arrows), and the presence of notable terms (ovals). Over time, *air* went from the core to the periphery while *acid* did the opposite, and this was accompanied by the removal of *calx*, *phlogiston*, and the emergence of *oxygen*.

a concept's lexemes or readings from core to periphery, and evaluates the recombination of these readings into new core concepts for onomasiological<sup>1</sup> assessment.

As a case study, we analyze two competing explanations for combustion during the Chemical Revolution: phlogiston theory (e.g., “a lighter burns because it contains phlogiston”) and oxygen theory (e.g., “a lighter burns because some substances ignite when in contact with oxygen”). Considering these theories as prototypical concepts (see Figure 1) with central and peripheral readings (Geeraerts, 1997), the overarching research question of this work is concerned with modeling onomasiological change in scientific discourse with temporal graphs. Our central question is whether the network structure reflects the spread and competition of readings in changing scientific concepts.

The rest of the paper is organized as follows.

<sup>1</sup>Onomasiological change refers to shifts in the linguistic expressions used to denote the same underlying concept. Instead of tracking how a word's meaning evolves (a semasiological perspective), the onomasiological approach fixes a concept and studies competition, replacement, or specialization among its lexical realizations over time (cf. Geeraerts (2009))

Section 2 motivates the study of concepts, especially scientific ones, and explains how prototype semantics serves as the theoretical framework for representing concepts as networks. Section 3 reviews related work on modeling concept formation using networks and introduces our task of detecting conceptual structure shifts. Section 4 describes the corpus and experiments, and Section 5 reports the results and their interpretation. Finally, Section 6 concludes and provides future research directions.

## 2 Background

This section begins by explaining why modeling concepts is important, especially in the scientific domain, and presents the linguistic theory from which we approach concepts as graphs. It ends with a brief historical background on the Chemical Revolution.

### 2.1 Concepts and Networks

**Why Concepts Matter** Concepts are mental tools designed to manage the (im)balance between expectation and experience (Koselleck, 2006). For instance, the concept of *nature* today is approached as an object rather than subject, which presents interpretations that diverge from those of previous societies. Since concepts are indicators of historical change, tracing their shifts is key to reconstruct how past societies understood and structured their world.

**Why Scientific Concepts** Unlike everyday concepts (e.g., *chair*), scientific concepts (e.g., *DNA*) cannot be formed by the subject through immediate experience of the object. They are learned through the academic establishment and lose meaning if isolated from this universal system of related ideas since they rely on neither moral judgments, gut feelings nor authority. This is how Vygotsky (1994) explains why scientific concepts are "true concepts" (viz., truly learned, nonspontaneous concepts). Since they are not impacted by cultural ambiguity, as study objects they enable "universal sensors" for historical change.

**Why Complex Networks** According to prototype semantics, concepts have readings or interpretations located at different levels; going from the center to the periphery, where the centroids are considered to be the most typical (Geeraerts, 1997). To illustrate this, consider how the concept of *fruit* generates a typical reading that varies depending

on the cultural perspective. This structure is akin to a graph's, and the literature supports that representing them as such offers means to model the meaning dynamics as well as social interactions that lead to concept formation (Kaye et al., 2024; Kedrick et al., 2024; Ju et al., 2020).

### 2.2 Chemical Revolution Overview

When science undergoes dramatic ontological changes (i.e., the whole system of concepts and their laws is replaced by a new one), by analogy with political uprisings, such processes are called scientific revolutions (Kuhn, 2012). The textbook example of this occurred in the 18th century, when competing sides disagreed sharply about what counted as an element and what was a compound, a foundational difference many chemists of the time had (Chang, 2015). Notably, Lavoisier's oxygen paradigm overtook the dominant one in chemistry for combustion, Stahl's phlogiston paradigm; while phlogistonists regarded metals as calx compounds (i.e., today known as oxides) plus phlogiston, oxygenists classified calxes as compounds of pure metals plus oxygen (i.e., today known as element), which cleared out research questions at the time related to not only combustion but calcination and respiration (Thagard, 1990), producing the ground work for many additional transformations in chemistry (Holmes, 1994). Arguably, phlogistonists' contributions were hindered by interpreting phenomena using phlogiston as a principle or "central reading", an ontological category that disappeared as Lavoisier's system developed (Chang, 2015).

Hence, this scientific scenario is optimal as a case study to model how concepts are changed, new ones arise, and old ones disappear. In particular, we focus on the decades where the new concept of oxygen became established while phlogiston vanished: from the 1750s to 1800s (Chang, 2011; Thagard, 1990). This scenario, however, only serves as a first step towards a generalizable approach to model conceptual transitions and the meaning dynamics involved in these changes.

## 3 Related Work

This section introduces the state-of-the-art methods for the NLP task most related to our problem (i.e., lexical-semantic change detection), the current challenge of improving interpretability, and how we approach it in this work as a conceptual structure shift detection task.

### 3.1 NLP Research Overview

The evolution of word meanings, often termed semantic change, has long been a central focus of linguistic inquiry. This process is crucial for understanding how language adapts to shifting cultural and communicative contexts. Advances in computational methods have provided novel tools to study semantic change at scale. Tahmasebi et al. (2021) and Periti and Montanelli (2024) offer comprehensive surveys on these methods, with a focus on diachronic word embeddings and neural methods. Diachronic embeddings capture the shifting meanings of words across time by mapping their usage in successive temporal intervals (Hamilton et al., 2016b; Dubossarsky et al., 2017). While such approaches effectively track semantic drift, they often fall short in allowing to understand the underlying mechanisms driving these changes.

Modeling evolving scholarship considering concept networks offers an alternative, more informative perspective. By building semantic graphs that organize a concept's core and peripheral readings, Kedrick et al. (2024) found conceptual structures with rigid cores hinder scientific innovation. They explain this using the network structure: the recombination of cores with fellow cores and peripheral entities is less likely to occur in disciplines where core concepts are more stable. Such framing is compatible with prototype semantics, which describes words as radial structures (e.g., word *fruit* with *apple* and *guava* as central and peripheral readings, respectively) where depending on the situational context some readings are more "central" or "typical" than others, boundaries are not rigid, and different concepts can overlap (Geeraerts, 1997).

Networks are thus informative in the sense that they can represent the meaning dynamics taking place alongside the social interactions driving them. By constructing community networks based on a topic model, Malaterre and Lareau (2023) showed the evolution of communities and their main research themes, tracing the development of the specialization areas that structure the field of philosophy of science today. Topic models are known for approaching documents as mixtures of abstract themes, where each theme is represented by a probability distribution over words (Blei et al., 2003). Document-topic distributions thus offer insight into how specific themes are distributed across time, and to our knowledge this approach was first implemented by Griffiths and Steyvers (2004) to find

scientific disciplines and evaluate collaborations between researchers.

### 3.2 Towards a Conceptual Structure Shift Detection Task

We approach concepts as prototypical structures whose meaning evolves through their relational embedding in domain-specific discourse. The conceptual structure shift detection task therefore frames conceptual change as a network-level phenomenon rather than a purely lexical one.

- **Input:** A diachronic scientific corpus (e.g., spanning multiple decades of publications from one or more research domains).
- **Output:** Concept networks for each time slice that capture core-periphery organization and their diachronic trajectories.

In the context of scientific revolutions, these trajectories should reflect conceptual structure shifts that we define as *periods where the prototypical core of a concept either reconfigures or expands towards new peripheries, resulting in multiple meaning shift cases*, namely

- the lexemes referring to the core diversify (i.e., **onomasiological change**),
- new overall senses for the core emerge (i.e., **semasiological change**),
- and/or a new ontological system replaces the previous one (i.e., **ontological change**).

None of these changes are exclusive and arguably happen in sequence (e.g., during the Chemical Revolution, onomasiological/semasiological changes led to the ontological one). Moreover, to explain these cases systems are expected to produce both static models per period and dynamic mappings across periods, **prioritizing interpretability in structural metrics**.

Hence, to alleviate interpretability concerns we estimate concepts as themes with a ranked vocabulary using topic modeling, then cluster the documents based on similarity. Doing so produces graphs with topic clusters that represent the concepts present in a decade, and how these concepts were structured in terms of core and peripheral words.

Regarding evaluation, this baseline adopts two established metrics from information theory to measure change:

- **Jensen-Shannon distance**, to quantify the difference across topic clusters by treating graphs as diachronic, referent-independent entities (i.e., a perspective looking from the future into the past and vice versa have a distance that is symmetrical).
- **Entropy**, to describe the magnitude of the change understood as "topic diversity" according to Hall et al. (2008).

In summary, we approach concepts as prototypical structures and define our task as conceptual structure shift detection: the input is a diachronic scientific corpus, and the expected output are concept networks that offer core and peripheral readings trajectories to inform the **onomasiological change assessment**, which is the case we decided to focus on for this work.

## 4 Material and Methods

Based on the Royal Society Corpus (RSC; Kermes et al. 2016; Fischer et al. 2020), this approach consisted of four stages: first, create a representative corpus by filtering publications given specified decades and oxygen terminology (Chang, 2011; Bizzoni et al., 2021). Second, represent concepts and their diachronic readings using topic modeling (Blei et al., 2003; Sievert and Shirley, 2014). Third, form concept networks by linking documents based on the Jensen-Shannon distance (Shannon, 1948) among topics and optimize connectivity by minimizing the percolation threshold (Radicchi, 2015); this is the tipping point where isolated nodes in a graph abruptly form the largest mutually reachable subgraph, with a low threshold indicating extensive connectivity using minimal resources. Finally, we reveal hidden groups by applying clustering (Lukasová, 1979) and community detection (Blondel et al., 2008) algorithms.

**Dataset Description** The RSC is a diachronic corpus of English scientific writing covering 47,837 publications and their metadata (e.g., publication year, author) in the Philosophical Transactions and Proceedings of the Royal Society of London from 1665 to 1996 (Fischer et al., 2020). We focus on 6 decades (1750-1800) to account for the events where oxygen developed as a concept; from 1774, when Joseph Priestley conformed to the theories of the time and introduced the gas as "dephlogisticated air" (Brown, 1997), till 1789, when Lavoisier presented the gas with a new term and led to the

widespread adoption of "oxygen" among chemists (Thagard, 1990). Doing so is supported by previous studies based on the RSC that in the same period report documents and lexico-grammatical changes addressing combustion, gases and chemical reactions, key themes in the phlogiston-to-oxygen conceptual transition (Degaetano-Ortlieb and Teich, 2019; Bizzoni et al., 2020; Teich et al., 2021).

### 4.1 Data Processing and Sampling

We retrieved texts from the RSC with a publication date between 1750–1800, defined decade batches to make our results comparable with previous studies on this corpus (Teich et al., 2021), then filtered documents to include instances of terms relevant to the discovery of oxygen. These terms were derived from Bizzoni et al. (2021), which used Kullback-Leibler Divergence (KLD) to identify distinctive terms during the Chemical Revolution; and Chang (2011), who explains why the term *oxygen* persisted despite theoretical revisions, arguing that its operational meaning remained stable while phlogiston theory lost operational grounding. Chang (2011) also notes that the theoretical meaning of phlogiston was conceptually close to later notions of energy, and describes phlogiston as a precursor to ideas such as free electrons.

After filtering the documents ( $N = 2337$ ), we performed two cleaning rounds. First, we removed special characters and terms shorter than two characters, converted the text to lowercase, and removed stopwords based on the standard list imported as "ENGLISH\_STOP\_WORDS" using scikit-learn (Buitinck et al., 2013); doing so substantially decreased the corpus size by removing prepositions, pronouns and articles (e.g., "a", "he" and "the") while retaining nouns, which according to Hamilton et al. (2016a) are more sensitive to linguistic changes. Regarding hyphenated terms, these were unchanged to preserve their semantic unity during topic modeling. Second, we filtered for nouns and adjectives, and lemmatized the corpus using tntoolkit<sup>2</sup>.

To test the influence of prior textual data on downstream analyses, for the six decades we implemented two sampling strategies: 1) cumulative sampling, where each decade includes the indicated decade and all previous decades' documents, and 2) non-cumulative sampling, where each decade includes only documents from that period.

<sup>2</sup><https://tntoolkit.readthedocs.io/en/latest/>

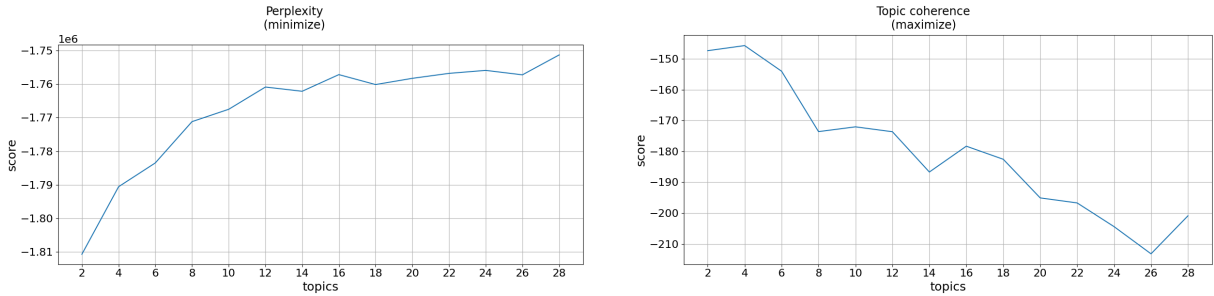


Figure 2: **Topic models evaluation.** These results, produced by evaluating models for the 1800s non-cumulative corpus, were consistent across decades and strategies. Since both metrics degrade beyond 6 topics ( $\uparrow$ perplexity,  $\downarrow$ coherence), the optimal number is 6.

## 4.2 Topic Modeling

In the context of the RSC, topic modeling has previously been used to estimate scientific disciplines (Fankhauser et al., 2016). The authors employed Latent Dirichlet Allocation (LDA), a probabilistic model designed to uncover abstract categories given a collection of documents (Blei et al., 2003), and found 30 distinct disciplines and sub-disciplines to describe the publications.

Our topic modeling workflow consisted of constructing a document-term matrix from the preprocessed corpus, counting the occurrences of each lemmatized term per document. We evaluated model performance using log-likelihood (Bengio et al., 2003) and topic coherence (Mimno et al., 2011); doing so mitigates the often manual and arbitrary parameter selection in LDA (Schmidt, 2012; Mohr and Bogdanov, 2013). Following best practices in interpretability, we generated topic labels using LDAvis (Sievert and Shirley, 2014), which are determined based on relevance scores that are more meaningful compared to word rankings (i.e., if *air* is the top-5 word yet its relevance score is higher compared to other terms, this will be the topic label). This is how we defined the optimal number of topics and the topic labels (i.e., the term that "names" the topic) across decades and strategies.

## 4.3 Network Construction

We considered documents as nodes and formed edges based on the Jensen-Shannon distance (Shannon, 1948) between document-topic distributions; given distributions  $p$  and  $q$ , distance  $d$  is given by:

$$d = \sqrt{\frac{D(p \parallel m) + D(q \parallel m)}{2}}$$

where  $m$  is the pointwise mean of  $p$  and  $q$  and  $D$  is the KLD. This way we formed adjacency matrices,

later binarized by optimizing connectivity based on the percolation threshold (Radicchi, 2015). A threshold was necessary to reduce the network's complexity by limiting the number of nodes and edges to only the most meaningful ones.

Hence, the conceptual structure was defined as follows: topic labels (Sievert and Shirley, 2014) "name" the concepts, while the topic-word distributions establish the central (i.e., top-10) and peripheral (i.e., top-50) terms.

## 4.4 Topic Aggregation

To find hidden groups, first we applied Hierarchical Agglomerative Clustering (HAC; Lukasová, 1979), a method that consists of measuring the distance between data points iteratively until forming groups with data points that exhibit higher similarity to one another compared to points in other groups. Each data point starts as its own cluster and merges with others until a number of groups is reached (i.e., 6 according to the optimal number of topics). For visualization purposes, we used cosine similarity as the metric for both HAC and UMAP (McInnes et al., 2018) dimensionality reduction approaches due of its reported performance in topic-based networks (Luhmann and Burghardt, 2022).

Similarly, the Louvain algorithm (Blondel et al., 2008) estimates groups by maximizing modularity, a metric that measures the quality of a network's division into distinct groups or communities, quantifying the density of intra- vs. inter-cluster connections. We used it as a complementary method to further analyze network dynamics among the topics (Malaterre and Lareau, 2023).

## 5 Results and Discussion

This section describes our findings regarding the phlogiston-to-oxygen conceptual transition: a topic cluster indicates onomasiological change given its



underplayed acidity as a secondary effect of phlogiston release (e.g., experiments burning sulfur produced a "gaseous calx" that was perceived as acidic since it lacked phlogiston), Lavoisier considered acidity as a core explanatory principle to reframe calxes as oxides and oxygen as an enabler of acid formation (e.g., sulfur is a combustible that when in contact with oxygen forms sulfur dioxide, which in turn dissolves in water to produce sulfurous acid). This is explicit in the results, where *calx* and *phlogiston* disappear from the periphery in the concept structure after *acid* takes over the core (cf. notable terms in Figure 1).

To compare the magnitude of diachronic change in both sampling strategies, we measured the topics "diversity of ideas" using entropy (Hall et al., 2008). Figure 4 presents different trends: while the cumulative strategy shows rising entropy across oxygen-related topics (illustrated by different colors), the non-cumulative declines then rises post-1774 oxygen discovery. This decline of diversity pre-1780 could be interpreted as the resistance of phlogistonists to de-center *air* in their findings (e.g., Priestley's "dephlogisticated air"), while the rise of diversity post-1780 could be attributed to Lavoisier coining the term *oxygen*. These observations support Thagard (1990)'s on why Lavoisier made the conceptual shift instead of Priestley. However, to clarify this, future studies will have to incorporate Lavoisier's articles in French, which are absent from our corpus.

## 5.2 Temporal Graphs

Considering the onomasiological change (Figure 3) and rising entropy (Figure 4), we built graphs based on the non-cumulative strategy; one for each decade, producing 6 in total (Figure A1).

To analyze network stability, we considered 5 parameters (Figure A2) from which follow three findings: first, although node size starts and ends similarly (400–500), edge density more than doubled (30 037–62 639). Second, declining network communities suggests integration (5–3), but falling modularity (0,39–0,19) disputes this: fewer, lower-quality communities indicate increased mixing. Lastly, percolation threshold rises in 1760s (0,3–0,54), falls in 1770s (0,54–0,32), then rises (0,32–0,58)—aligning with modularity.

Low threshold/high modularity signals efficient connectivity given well-defined communities (e.g., communication is faster when people are organized in groups and allotted resources are suffi-

cient), while high threshold/low modularity signals mixed communities requiring high connectivity effort (e.g., mixed groups hinder person-finding, requiring more channels and impeding communication).

## 6 Conclusion and Future Work

This work case-studies conceptual change during the Chemical Revolution (i.e., phlogiston-to-oxygen) via interpretable complex networks. Our analysis reveals two findings: 1) a topic cluster shifts from *air* to *acid*, showing onomasiological change given the movement of core/peripheral readings accompanied by higher entropy, and 2) network communities and modularity decline, while percolation and edge density rise, indicating higher connectivity effort. These results indicate that changing concepts form high entropy clusters (i.e., more variability of ideas) that increase topological density in the knowledge network (i.e., more effort is needed for community finding and message passing), potentially hindering communication. Future work will refine representation of rare terms and add directed edges (e.g., colexification, citations linked to terms) to account for semantic and social directionality, and predict inter-temporal graph links via node/edge embeddings to identify influential documents and authors.

## Limitations

Our corpus so far is limited to English scientific texts, which does not account for the whole spectrum of ideas exchanged during the Chemical Revolution. Adding multilingual and multi-genre texts is a desiderata to work towards a comprehensive picture. In this study, we derived core and peripheral terms from topic-word distributions rather than cluster centroids, potentially overlooking document embedding geometries. Rare peripheral terms receive sparse representations, potentially missing subtle shifts. As an approach that builds undirected graphs, the formed edges overlook semantic directionality; representing this is important to understand how concepts evolve (e.g., through meaning gain or loss), making of the conceptual structure a phylogenetic tree (Carling et al., 2023). Future work will address these challenges via directed graphs and rare-term augmentation, as well as the estimation of phylogenetic structures. Moreover, to test the generalizability of the approach, first steps will include modeling conceptual transitions

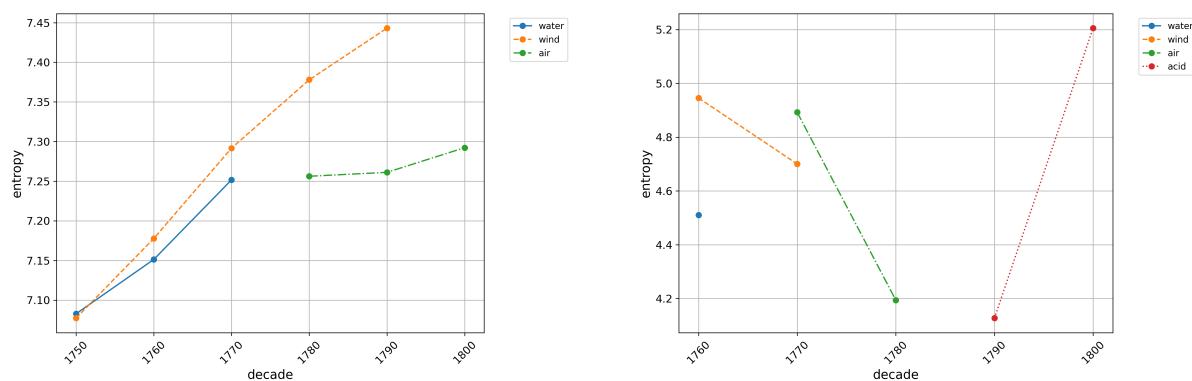


Figure 4: **Topic entropy over time.** While in cumulative sampling (left) entropy trajectories of oxygen-related terms increase over time, non-cumulative sampling (right) shows a decline then rise during the development of oxygen theory (cf. 1772-1789 in Thagard (1990)).

across disciplines (e.g., germ plasm to chromosomal genes in biology).

## Acknowledgments

Funded by the European Union under grant 101119511. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

## References

- Christin Beck. 2024. [Review of Tahmasebi, Borin, Jantowt, Xu & Hengchen \(2021\): Computational Approaches to Semantic Change](#). *Journal of Historical Linguistics*, 14(2):376–384.
- Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. *J. Mach. Learn. Res.*, 3:1137–1155.
- Yuri Bizzoni, Stefania Degaetano-Ortlieb, Peter Fankhauser, and Elke Teich. 2020. [Linguistic Variation and Change in 250 Years of English Scientific Writing: A Data-Driven Approach](#). *Frontiers in Artificial Intelligence*, 3.
- Yuri Bizzoni, Stefania Degaetano-Ortlieb, Katrin Menzel, and Elke Teich. 2021. [The diffusion of scientific terms – tracing individuals’ influence in the history of science for English](#). In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 120–127, Punta Cana, Dominican Republic (online). Association for Computational Linguistics.
- Andreas Blank. 2013. [Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change](#). In Andreas Blank and Peter Koch, editors, *Historical Semantics and Cognition*, pages 61–90. De Gruyter Mouton.
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.
- Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. [Fast unfolding of communities in large networks](#). *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- Lee K. Brown. 1997. [“Dephlogisticated Air” Revisited: Oxygen Treatment for Central Sleep Apnea Syndrome](#). *CHEST*, 111(2):269–271. Publisher: Elsevier.
- Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. 2013. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD workshop: Languages for data mining and machine learning*, pages 108–122.
- Gerd Carling, Sandra Cronhamn, Olof Lundgren, Victor Bogren Svensson, and Johan Frid. 2023. [The evolution of lexical semantics dynamics, directionality, and drift](#). *Frontiers in Communication*, 8.
- Hasok Chang. 2011. [The Persistence of Epistemic Objects Through Scientific Change](#). *Erkenntnis*, 75(3):413–429.
- Hasok Chang. 2015. [The Chemical Revolution revisited](#). *Studies in History and Philosophy of Science Part A*, 49:91–98.
- Stefania Degaetano-Ortlieb and Elke Teich. 2019. [Toward an optimal code for communication: The case of scientific English](#). *Corpus Linguistics and Linguistic Theory*, 18(1):175–207.
- Haim Dubossarsky, Daphna Weinsall, and Eitan Grossman. 2017. [Outta control: Laws of semantic change and inherent biases in word representation models](#).



- In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145, Copenhagen, Denmark. Association for Computational Linguistics.
- Peter Fankhauser, Jörg Knappen, and Elke Teich. 2016. [Topical Diversification Over Time In The Royal Society Corpus](#).
- Stefan Fischer, Jörg Knappen, Katrin Menzel, and Elke Teich. 2020. [The Royal Society Corpus 6.0: Providing 300+ Years of Scientific Writing for Humanistic Study](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 794–802, Marseille, France. European Language Resources Association.
- Dirk Geeraerts. 1997. *Diachronic Prototype Semantics: A Contribution Historical Lexicology*. Oxford University Press.
- Dirk Geeraerts. 2009. *Theories of Lexical Semantics*. Oxford University Press.
- Stefan Gries and Martin Hilpert. 2008. [The identification of stages in diachronic data: Variability-based neighbour clustering](#). *Corpora*, 3:59–81.
- Thomas L. Griffiths and Mark Steyvers. 2004. [Finding scientific topics](#). *Proceedings of the National Academy of Sciences*, 101(suppl\_1):5228–5235. Publisher: Proceedings of the National Academy of Sciences.
- David Hall, Daniel Jurafsky, and Christopher D. Manning. 2008. Studying the History of Ideas Using Topic Models. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 363–371, Honolulu, Hawaii. Association for Computational Linguistics.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016a. [Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2116–2121, Austin, Texas. Association for Computational Linguistics.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016b. [Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.
- Frederic L. Holmes. 1994. What was the chemical revolution about. In *Abstracts of Papers of the American Chemical Society*, volume 208, pages 13–HIST, USA. American Chemical Society.
- Harang Ju, Dale Zhou, Ann Sizemore Blevins, David M. Lydon-Staley, Judith Kaplan, Julio Roberto Tuma, and Danielle S Bassett. 2020. [The network structure of scientific revolutions](#).
- Aleksandra Kaye, Raphael Schlattmann, Malte Vogl, Bernardo S. Buarque, Jascha Schmitz, Lea Weiß, and Laura von Welczeck. 2024. Socio-Epistemic Networks: A Framework for History of Knowledge. In Charlotte A. Lerg, Johan Östling, Jana Weiß, Anne Kwaschik, and Claudia Roesch, editors, *History of Intellectual Culture 3/2024: Experimental Spaces: Knowledge Production and Its Environments in the Long Nineteenth Century*, pages 253–274. De Gruyter Oldenbourg.
- Kara Kedrick, Ekaterina Levitskaya, and Russell J. Funk. 2024. [Conceptual structure and the growth of scientific knowledge](#). *Nature Human Behaviour*, 8(10):1915–1923. Publisher: Nature Publishing Group.
- Hannah Kermes, Stefania Degaetano-Ortlieb, Ashraf Khamis, Jörg Knappen, and Elke Teich. 2016. [The Royal Society Corpus: From Uncharted Data to Corpus](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1928–1931, Portorož, Slovenia. European Language Resources Association (ELRA).
- Reinhart Koselleck. 2006. Die Verzeitlichung der Begriffe. In Ulrike Spree and Willibald Steinmetz, editors, *Begriffsgeschichten: Studien Zur Semantik Und Pragmatik Der Politischen Und Sozialen Sprache*, Suhrkamp Taschenbuch Wissenschaft, page 85. Suhrkamp, Frankfurt.
- Thomas Kuhn. 2012. *The Structure of Scientific Revolutions: 50th Anniversary Edition*. University of Chicago Press, Chicago, IL.
- Jan Luhmann and Manuel Burghardt. 2022. [Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape](#). *Journal of the Association for Information Science and Technology*, 73(2):148–171.
- Alena Lukasová. 1979. [Hierarchical agglomerative clustering procedure](#). *Pattern Recognition*, 11(5-6):365–381. Publisher: Elsevier.
- Christophe Malaterre and Francis Lareau. 2023. [Identifying Hidden Communities of Interest with Topic-based Networks: A Case Study of the Community of Philosophers of Science \(1930-2017\)](#). In *Hawaii International Conference on System Sciences*.
- Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. [UMAP: Uniform manifold approximation and projection](#). *Journal of Open Source Software*, 3(29):861.
- David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11*, pages 262–272, USA. Association for Computational Linguistics.

- John W. Mohr and Petko Bogdanov. 2013. [Introduction—Topic models: What they are and why they matter](#). *Poetics*, 41(6):545–569.
- Francesco Periti and Stefano Montanelli. 2024. [Lexical Semantic Change through Large Language Models: a Survey](#). *ACM Comput. Surv.*, 56(11):282:1–282:38.
- Filippo Radicchi. 2015. [Predicting percolation thresholds in networks](#). *Physical Review E*, 91(1):010801.
- Benjamin M Schmidt. 2012. [Words Alone: Dismantling Topic Models in the Humanities](#). *Journal of Digital Humanities*, 2(1):49–65.
- Claude Elwood Shannon. 1948. [A mathematical theory of communication](#). *The Bell System Technical Journal*.
- Carson Sievert and Kenneth Shirley. 2014. [LDAvis: A method for visualizing and interpreting topics](#). In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70.
- Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen. 2021. [Computational approaches to semantic change](#). Language Science Press, Berlin.
- Elke Teich, Peter Fankhauser, Stefania Degaetano-Ortlieb, and Yuri Bizzoni. 2021. [Less is More/More Diverse: On The Communicative Utility of Linguistic Conventionalization](#). *Frontiers in Communication*, 5.
- Paul Thagard. 1990. [The Conceptual Structure of the Chemical Revolution](#). *Philosophy of Science*, 57(2):183–209. Publisher: [The University of Chicago Press, Philosophy of Science Association].
- Lev Semenovič Vygotsky. 1994. The Development of Thinking and Concept Formation in Adolescence. In René van der Veer, editor, *Vygotsky Reader*, pages 185–266. Blackwell, Oxford.
- Ladislav Zgusta. 2011. Onomasiological change: Sachen-change reflected by Wörter. In Edgar C. Polomé, editor, *Research Guide on Language Change*, pages 389–398. De Gruyter Mouton.

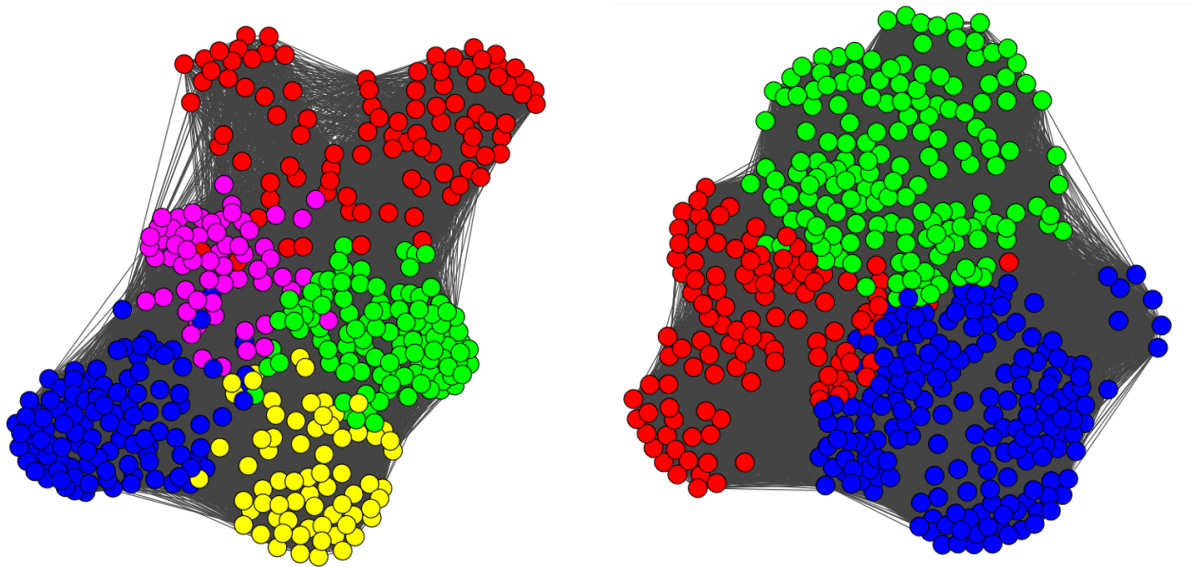


Figure A1: **Temporal graphs.** The colors illustrate the number of communities which went on decline: starting at 5 (1750s) and ending at 3 (1800s).



Figure A2: **Network metrics.** We used five parameters to interpret network stability over time: nodes size, edge density (where  $y = 1 \times 10^3$ , e.g., last number of edges reported is >60 000), communities count, modularity and percolation threshold.