Discursive parallels of the chemical revolution
Topic modelling and distributional analysis

Sofía Aguilar-Valdez, Bách Phan-Tất, Dirk Speelman, Dirk Geeraerts, and Stefania Degaetano-Ortlieb

1. Introduction

How do conceptual changes show up in scientific writing? This paper addresses this question by examining both the topics that authors foreground and the immediate linguistic contexts in which key terms are embedded. Focusing on the late phase of the Chemical Revolution (1750–1800), we analyze a subset of the Royal Society Corpus (Kermes et al., 2016; Fischer et al., 2020). We adopt a two-level approach that connects broad thematic trends at the document level with local linguistic evidence at the word level, allowing us to link where change appears in the discourse to how it is linguistically realized.

At the document level, we trace how topics emerge, stabilize or decline across decades. At the word level, we focus on the concept of *air*, a central term in late eighteenth-century chemistry, and investigate how it undergoes conceptual change by analyzing the company it keeps, following Firth's (1957) distributional insight. While our approach is grounded in distributional semantics, a perspective that also underlies LLMs, we deliberately rely on more established methods of topic modelling based on Latent Dirichlet Allocation (LDA; Blei et al., 2003) combined with interpretable information-theoretic measures. This choice is not merely pragmatic. For questions in the field of HPSS, it is essential to make explicit how epistemic knowledge is structured, contested, and consolidated in discourse. Probabilistic topic models yield explicit probability distributions over themes, which can then be analyzed using information-theoretic measures based on entropy or divergence. In this framework, entropy measures the diversity of topics within a period. Rising entropy signals periods in which multiple thematic perspectives coexist, corresponding to phases of epistemic competition, while declining entropy indicates consolidation around fewer, more dominant themes. Although LLMs and embedding-based approaches can also quantify semantic shifts, such divergence is defined over opaque latent representations that encompass multiple underlying sources of variation (lexical, semantic, grammatical, etc.), which cannot be easily disentangled, making the interpretability considerably more difficult

(Tahmasebi et al., 2021). In particular, these approaches do not provide direct access to the grammatical configurations in which a concept is embedded, nor do they allow changes in syntactic function and lexical realization to be inspected independently. Data sparsity in historical corpora further reinforces this methodological choice (see below and cf. Lareau and Malaterre, 2026). We do, however, also sketch ideas for taking the analysis presented here to the next level, which would include the generation of a graph structure of concepts which could then be enriched with an LLM-based analysis. By way of presenting our analysis, we also intend to familiarize the reader with key concepts for the computational analysis of semantic change—be it by LLMs or other means.

Our document-level analysis asks a simple question: which topics are most characteristic of each decade, and how much do those decade profiles differ from one another? The results reveal a recognizable sequence tied to oxygen-centered chemistry, with decades in which themes around *water*, then *air*, and later *acid* gain comparative prominence.

Our word-level view zooms in on *air*, and quantifies its changing use by examining the words that most often appear in its immediate context—especially adjectives such as *fixed*, *dephlogisticated*, and *vital*. By grouping surrounding words into syntactic "slots" (for example, adjectives that modify *air*), we track how the distribution of slot fillers changes over time. This makes recontextualization visible as a measurable shift from general or environmental descriptors toward chemically specific ones.

Taken together, these two perspectives—topic trends across documents and local context around air—converge on the same picture: a gradual reorganization of discourse rather than a single abrupt break. This interpretation is compatible with standard histories of the oxygen program, while remaining grounded in observable textual patterns (cf. Holmes, 1994; Thagard, 1990; and Bensaude-Vincent and Abbri, 1995 from a discursive perspective). Building on earlier work showing that lexical and grammatical distributions mark periods of scientific reorientation (Degaetano-Ortlieb and Teich, 2018), our finer-grained slot-filler analysis complements document-level trends by identifying the local linguistic contexts in which conceptual change is realized. This aligns with usage-based and variationist perspectives in lexical semantics, where changes in distribution are meaningful for how concepts are encoded (Geeraerts, 2009; Geeraerts, 1997), and ties directly to our empirical observations on *air* in the late eighteenth century.

Finally, the two-level signals (document-level topics and local linguistic evidence) identified here will serve as input in future graph-based modeling of conceptual change.

2. Related work

From a cognitive linguistic perspective, Geeraerts (2009) and Geeraerts et al. (2012) demonstrate how meanings evolve around culturally salient prototypes that shift over time. For instance, in the category bird, a robin or sparrow is often considered more prototypical than a penguin or ostrich, because it more closely matches the characteristics people associate with birds, such as flying or singing. However, these prototypes can shift over time and across communities. Geeraerts et al. (2012) further highlight how different communities may use the same words differently, depending on their social or disciplinary context. The idea that changes in term usage or terminology often reflect shifts in conceptual framing has also been explored in diachronic corpus linguistics. Before a shift, competing perspectives often surface and circulate within a scientific community. Once the shift occurs, familiar terms may acquire new meanings, or entirely new terminology may emerge to capture novel concepts. Degaetano-Ortlieb and Teich (2019) show that during revolutionary periods, the lexicon exhibits a wave-like tendency in which new terms are coined and old terms acquire new meanings or are used in different linguistic environments (e.g. in different syntactic positions, i.e. as a subject or object). These shifts often precede or co-occur with major scientific reorientations. After a shift the diverging peaks within the lexicon stabilize showing periods of consolidation.

An illustrative example is the chemical revolution of the late 18th century, which centered around the topic of combustion. At the time, the dominant explanation for combustion was the phlogiston theory, which stated that a fire-like substance called phlogiston was released during burning. However, increasingly detailed experimental work, especially by Lavoisier and Priestley, led to the discovery of oxygen, a new interpretation of combustion as a process of oxidation. This finding posed a fundamental anomaly that the phlogiston paradigm could not explain. The in-depth experimental work on combustion within the phlogiston paradigm led to the oxygen discovery, which eventually became the established paradigm (Kuhn, 1962, pp. 66–76; 148–150). In this case, a new framework replaced an older one, and the scientific field moved toward a paradigm that offered a different perspective on both old and new problems.

Recent work in computational semantics has greatly advanced the analysis of semantic change. Surveys by Tahmasebi et al. (2021) and Cassotti et al. (2024) categorize methods into frequency-based, context-based, and embedding-based approaches, emphasizing challenges in aligning models across time periods. Periti et al. (2022) address this challenge through incremental updating of word embeddings, which facilitates continuous tracking of word usage shifts without the need for post hoc alignment. Their approach has been applied to diachronic corpora in both scientific and political domains.

In their recent review, Periti and Montanelli (2024) discuss how large language models (LLMs) can be leveraged for semantic shift detection. While LLMs capture nuanced contextual meanings, their lack of transparency limits interpretability. Periti and Tahmasebi (2024) introduce evaluation protocols for comparing LLM-based shift detection with classical distributional approaches, showing that LLMs outperform in capturing nuanced sense distinctions but tend to overfit when applied to sparse historical data, which is exactly the case in our study.

In this study we prioritized interpretability with a framework based on topic modeling and information-theoretic metrics, aiming to explore the conceptual shifts around the chemical revolution from two linguistic dimensions: semantics, expecting distinct themes over time and drastic change in their dynamics after the discovery of oxygen, and lexico-grammar, where syntactic relationships (slots) and their lexical content (slot fillers) would reflect change that would be consistent with the literature (Degaetano-Ortlieb and Teich, 2019; Chang, 2011). In terms of methods, we combine document-level topic modeling with interpretable syntactic slot-filler analysis. While most studies focus on either word-level or topic-level analysis, we integrate both to bridge the gap between lexical semantic drift and conceptual reorganization. We test this approach envisioning it to offer a richer, more interpretable model of how linguistic traces reflect underlying changes during periods of change.

3. Methods

This section describes the dataset and why it is well-suited for our task, and the methods used to represent and analyze the linguistic dimensions of a conceptual shift. The code[1], data samples, and full-texts[2] are available. As a case study, our aim is to analyze two competing theories during the period of 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"). We seek to explore the decline and rise of phlogiston and oxygen respectively through two dimensions: at the **document level**, we estimate conceptual properties of the drift, at the **word level**, we aim to represent linguistic properties of the drift. By combining these approaches, our goal is to explore to what extent these reflect the conceptual change produced by the discovery of oxygen.

3.1. Data

We draw on the Royal Society Corpus (RSC), a diachronic corpus of English scientific writing covering publications in the Philosophical Transactions and Proceedings of the Royal Society of London from 1665 to 1996 (Kermes et al., 2016; Fischer et al., 2020). The RSC is designed to support research on the development of scientific English, amounting at 47 837 texts. It provides metadata such as publication year, author, and document type, allowing for fine-grained analyses across linguistic and extralinguistic variables (Fischer et al., 2020; Menzel et al., 2021). We focus on the period between 1750 and 1800, which corresponds to the period before and after the Chemical Revolution. This period is particularly rich in documents addressing combustion, gases, and chemical reactions, key themes in the shift from phlogiston to oxygen theory. Preprocessing involved tokenization, lemmatization, and sentence splitting using spaCy. Lemmatized tokens were used throughout to enhance consistency and reduce lexical sparsity in our analyses.

3.2 Document level

This approach consisted of three stages. First, to create a representative corpus, we filtered publications given specified decades and oxygen terminology. Next, we estimated the most distinctive topics over time using topic modeling. Finally, to approach measurement of the shift assessment, we used the topic entropy trajectories to identify key topics and calculated the Jenson-Shannon (JS) distance between them.

1 https://github.com/sa-aguilarv/llms-hpss

2 https://fedora.clarin-d.uni-saarland.de/rsc_v6/

I. Data sampling We retrieved texts published in the period, then filtered those with oxygen terms derived from previous studies: 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. Lastly, to test prior text influence on downstream analyses, we created cumulative and non-cumulative datasets: in cumulative sampling each decade includes the indicated decade and all previous decades' documents, in non-cumulative sampling each decade includes only documents from that period. This is to test the influence of prior textual data on downstream analyses, i.e. how utilizing or not documents from the past would impact subsequent analyses.

II. Topic modeling Document-topic distributions offer insight into how specific themes are distributed across time, see e.g. Griffiths and Steyvers (2004) who applied this to find scientific disciplines and based on these evaluated collaborations between researchers. Our topic modeling workflow consisted of constructing a document-term matrix (DTM) from the preprocessed corpus, counting the occurrences of each lemmatized term per document. We evaluated model performance and topic interpretability using log-likelihood (Bengio et al., 2003) and topic coherence (Mimno et al., 2011), to mitigate 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).[3] We modeled topics separately for each sampling strategy (cumulative and non-cumulative) to evaluate the robustness of the trajectories. We found 2,337 documents after sampling, and that a model of 6 topics seems adequate as it achieved low perplexity and high coherence across all decades.

III. Shift assessment. Entropy, originally defined by Shannon (1948), quantifies the degree of unpredictability or diversity in a distribution. In our context, it measures the variability of each topic's probability distribution over time, indicating how conceptually stable or unstable a topic is. This approach has been used to capture thematic volatility in scientific discourse (Hall et al., 2008; Luhmann and Burghardt, 2022). We focused on topics whose top terms included one or more oxygen-related keywords and whose entropy trajectories showed a sequential activation across the decades. This allowed us to identify topics that plausibly represent distinct stages in the development of oxygen theory.

3 Note that, while current state-of-the-art approaches for scientific document classification rely on contrastive learning over citation graphs (Singh et al., 2023), this was not feasible for our study due to the RSC's sparse and non-standard citation formatting. We thus selected LDA as an interpretable and well-established method for estimating thematic shifts.

3.3 Word level

We used data from 1750 to 1799, filtered by the topic labels *Biochemistry*, *Chemistry 1*, and *Chemistry 2*, which were derived from an earlier LDA-based topic modeling study on the Royal Society Corpus (Fankhauser et al., 2016).

To analyze linguistic change at the micro-level, we investigated the lexeme *air*, a term central to both phlogiston and oxygen theories. The corpus was split into five-year time bins to capture fine-grained temporal resolution. All data were part-of-speech and dependency parsed using the Stanza library (Qi et al., 2020).

To examine syntactic and semantic shifts of *air*, we employed SynFlow (Phan-Tất, 2025), a custom Python package designed for extracting syntactic co-occurences of words from dependency-parsed corpora for collocational analyses. SynFlow builds a syntactic tree based on CONLLU format annotations and retrieves (and counts) the syntactic slots and slot fillers for a given target word or part of speech. For example, given the target *air* as a *NOUN*, some extracted slots include chi_det (determiner, as in *the air*) and pa_nsubj (subject, as in *the air heats up*). SynFlow uses the prefix chi (children) when the target is the head in a dependency relation and pa (parent) when it is the dependent. This distinguishes the directionality of syntactic relations. It also enables the reconstruction of syntactic slot-filler distributions. Thus, it was chosen for its ability to systematically retrieve and quantify the frequencies of both syntactic slots (e.g., chi_amod) and their fillers (e.g., adjectives like "pure" or "vital") across historical stages. In our workflow, we first used SynFlow to retrieve the frequencies of all syntactic slots of *air* in each five-year time bin (e.g., 1750–1755, 1755–1760) before calculating Jensen-Shannon Divergence (JSD) between slot distributions to measure distributional shifts over time. JSD is a symmetric and bounded measure that quantifies the difference between two probability distributions.

4. Results and discussion

We begin this section with a macro-analytical view by modeling the document level, analyzing topic drifts, and then move on to a micro-analysis at the word level by analyzing linguistic properties of the changes observed.

4.1 Topic drifts at document level

4.1.1 Topic trajectories across decades under cumulative and non-cumulative modeling

Table 1 shows the 6 topics and 3 top-words of every decade based on the cumulative strategy (decades and past decades), and the results suggest this strategy provides stable topics over time, with recurring labels such as *author*, *year*, *observation*, *distance*, *plant*, *water*, *air*, *wind*, *foot*, and *blood*. Note that topics sharing a column with different labels are not equivalent; the alignment is arbitrary.

Table 2 presents a seemingly more diverse set of topics through time. This is based on the non-cumulative strategy (only documents from that decade), and the results hint toward more fine-grained representations of the themes discussed. Here, the term *air* does not appear at all in the 1750s; it enters the topic of *water* in the 1760s, becomes a distinct topic label in the 1770s, and transitions into a top term under the *acid* topic in the 1790s. By the 1800s, *air* disappears as a salient term, leaving *acid* as the dominant theme.

Table 1: Topic labels of the 6 topics with their 3 top-words over time based on the cumulative strategy

	1		2		3		4		5		6	
		year		time		plant		water		water		blood
		author		observatio		fig		experimen		time		time
1750	**author**	time	**observation**	motion	**plant**	tree	**water**	air	**wind**	place	**blood**	body
		year		time		plant		water		time		blood
		author		observatio		fig		experimen		water		time
1760	**author**	number	**observation**	sun	**plant**	animal	**water**	air	**wind**	place	**blood**	body
		year		time		plant		water		wind		blood
		author		observatio		animal		air		water		time
1770	**year**	number	**observation**	distance	**plant**	fig	**water**	experimen	**wind**	time	**blood**	body
		year		time		plant		time		air		blood
		author		observatio		animal		water		water		body
1780	**year**	number	**observation**	distance	**plant**	fig	**air**	wind	**wind**	experimen	**blood**	time
		year		time		plant		air		time		blood
		author		observatio		fig		water		day		body
1790	**year**	number	**observation**	distance	**plant**	tree	**air**	experimen	**wind**	water	**blood**	animal
		year		time		plant		air		time		animal
		author		observatio		tree		water		foot		blood
1800	**year**	number	**distance**	distance	**plant**	fig	**air**	experimen	**foot**	water	**blood**	body

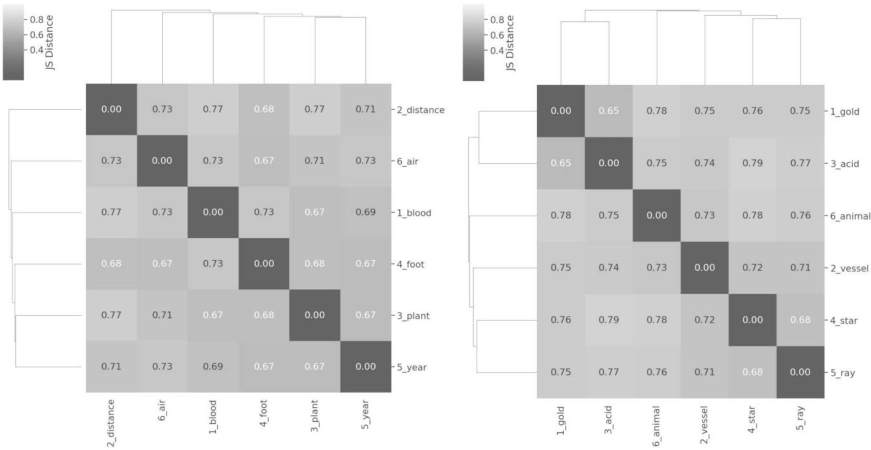
Table 2: Topic labels of the 6 topics with their 3 top-words over time based on the non-cumulative strategy.

	1		2		3		4		5		6	
		water		day		plant		time		number		foot
		experiment		time		animal		water		year		inscription
1750	**experiment**	glass	**day**	year	**plant**	body	**earthquake**	foot	**number**	time	**inscription**	word
		point		animal		time		wind		year		water
		line		body		observation		day		inscription		air
1760	**point**	ratio	**animal**	part	**observation**	sun	**wind**	foot	**year**	letter	**water**	experiment
		experiment		year		bird		observation		number		air
		body		animal		country		wind		height		water
1770	**ball**	inch	**animal**	tree	**bird**	water	**wind**	time	**number**	difference	**air**	experiment
		body		animal		star		day		degree		air
		inch		part		distance		time		thermometer		water
1780	**body**	velocity	**animal**	time	**star**	observation	**day**	foot	**thermometer**	experiment	**air**	acid
		experiment		eye		point		day		observation		acid
		inch		animal		line		time		star		air
1790	**inch**	water	**eye**	experiment	**gravity**	gravity	**day**	year	**observation**	time	**acid**	water
		experiment		animal		star		ray		iron		acid
		body		stomach		motion		glass		gold		water
1800	**vessel**	vessel	**animal**	portion	**star**	observation	**ray**	experiment	**gold**	copper	**acid**	gas

4.1.2 Distinctiveness of topics across decades

To assess the distinctiveness of topics across decades, we computed the JS distance between them, validating whether each decade had estimated distinct themes. Figure 1 shows a comparison of the 1800s: compared to the non-cumulative (decade without past decades) approach, the cumulative (decade including past decades) approach estimates topics that have lower scores which hints toward higher similarity. These results were consistent in all decades.

Fig. 1: JS distance scores for the 6 topics from the 1800s. From left to right, it shows the cumulative and non-cumulative strategies.



4.1.3 Assessing shifts in topics variability

To assess shifts in the variability of each topic's probability distribution over time, we used entropy as an indicator of how conceptually stable or unstable a topic is. We found higher values in the cumulative strategy (decade with past decades). Figure 2 shows topics with oxygen terms; notably, *wind* and *water* seem to behave similarly from the 1750s to the 1770s, while *air* seems to be the continuation of *water*. This is consistent with Tables 1 and 2, where the *air* term often appears with the water label and vice versa, supporting the view that *air* and *water* are thematically linked across decades. Based on the non-cumulative strategy (decade without the past), Figure 3 hints the trajectory of *wind* is continued by *air*, and from there *acid* follows. Overall, while the cumulative strategy shows rising entropy across oxygen-related topics, 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 decenter *air* in their findings (e.g., Priestley's *dephlogisticated air*), while the rise of diversity post-1780 could be attributed to Lavoisier's coining the term *oxygen*. These observations support Thagard's (1990) 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.

Lastly, based on the observed patterns, we selected representative topics given the following: (1) the non-cumulative strategy (decade without past) offers more fine-grained topics, (2) the topic of *water* seems to develop into the topic of *air*, and (3) the topic of *air* drifts to become the topic of *acid*. Figure 4 shows the JS distance between these topics, with two maximum values obtained between the 1760s-1800s (*water-acid*), and the 1780s-1800s (*air-acid*). Interestingly, these values seem to match the conceptual shift if we assume the *water* topic as reflecting phlogiston theory, *air* as a transition phase, and *acid* as representing the oxygen paradigm. In the 1770s, Priestley contested phlogiston theory by demonstrating the existence of what he called *dephlogisticated air,* a phenomenon that Lavoisier later conceptualized coining the term of oxygen (Chang, 2011). Hence, the key patterns of text reuse and rhetoric related to the competing oxygen and phlogiston concepts may be contained in the documents that are members of the topics *acid* (1800s) and *water* (1760s), respectively, while *air* (1780s) could be studied as a transition stage between both.

Fig. 2: Entropy values over time based on the cumulative strategy for oxygen-related terms.

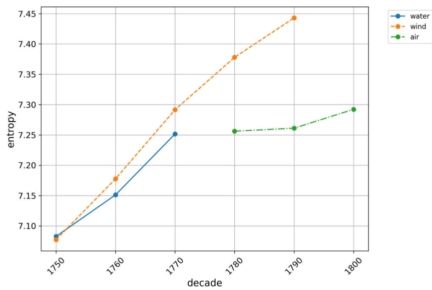


Fig. 3: Entropy values over time based on the non-cumulative strategy for oxygen-related terms.

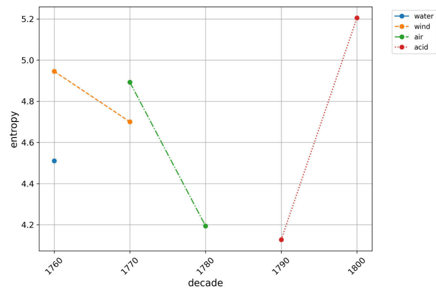
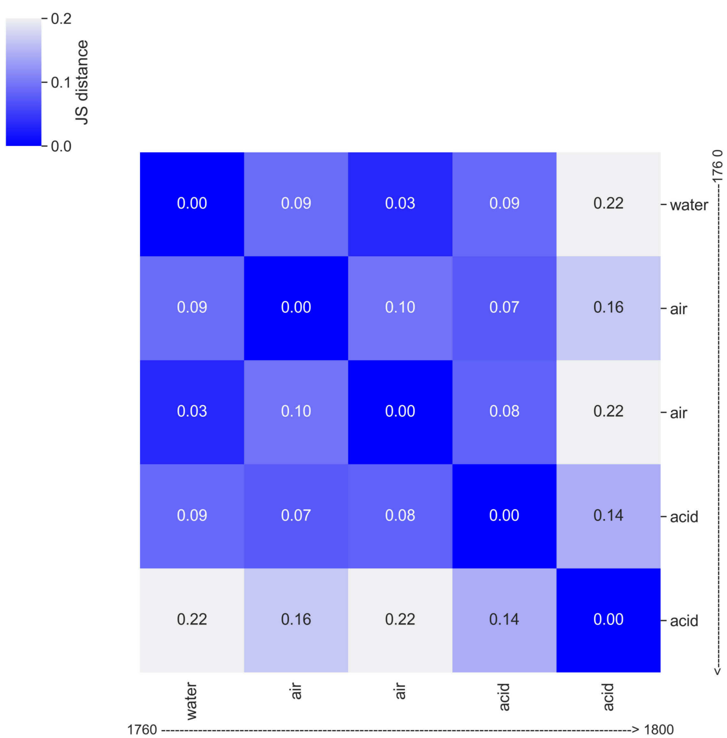


Fig. 4: JS distance across selected topics from the non-cumulative strategy: water (1760), air (1770), air (1780), acid (1790), and acid (1800).



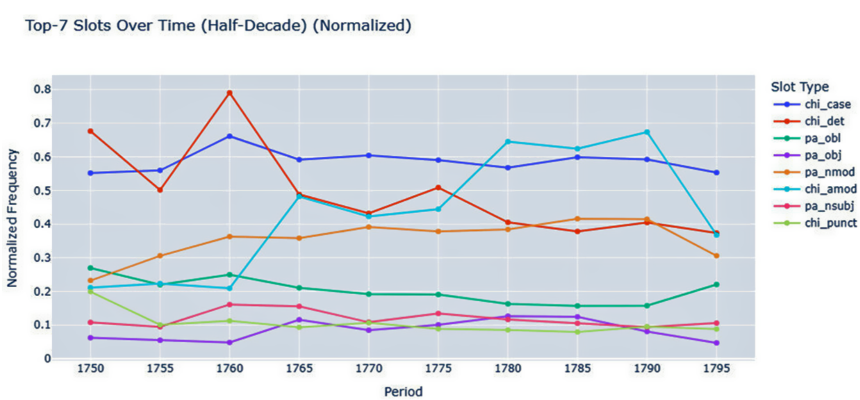
4.2 Word level

To investigate change at the word level, we focused on the transition phase and analyzed the term *air*, which plays a central role in the transition from phlogiston to oxygen theory as shown in our document level analysis. Using SynFlow, a syntactic analysis tool for dependency-parsed corpora, we extracted and analyzed the slot-filler distributions of *air* across five-year time bins between 1750 and 1800 as described in our method section. Our goal was to track how the linguistic environments of *air* evolved over time, as a proxy for changing conceptual and epistemic roles.

4.2.1 Diachronic distribution of syntactic slots associated with *air*

Figure 5 presents the relative frequencies of all syntactic slots associated with *air* over time. *Air* was syntactically stable as the relative frequency of the syntactic slots did not fluctuate much. Slots such as chi_obj (object) and pa_nsubj (nominal subject) remained stable in relative frequency throughout the period, indicating persistent grammatical roles. In contrast, the relative frequency of the chi_amod (adjectival modifier) and chi_det (determiner) slots showed quite some variation. The frequency of *amod* was stable at first, then suddenly increased during the first half of the 1760s, then further increased from the latter half of the 1760s to early 1790s then suddenly decreased. The frequency of *det* fluctuated at first then decreased gradually. This suggests a shift in how *air* was conceptualized. In earlier decades, *air* was frequently introduced as a known, observable entity (*the air*), consistent with a referential and phenomenological view. Over time, the decline of this pattern may reflect a growing interest in the internal properties of *air*, i.e. its quality, typology, and theoretical framing, rather than its immediate physical instantiation. This change aligns with the transition from descriptive to theoretical discourse in chemistry and also explains the increased use paired with a premodifying adjective (amod).

Fig. 5: Frequency of top-7 slots between periods



4.2.2 Syntactic signatures of the conceptual change of *air*

Figure 6: Jensen-Shannon Divergence of all slots

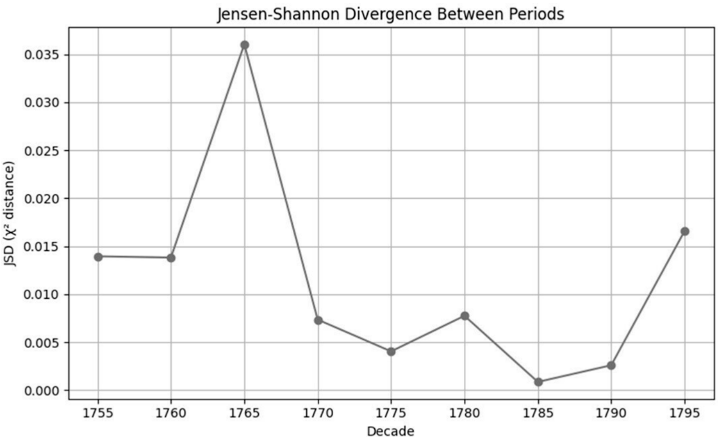
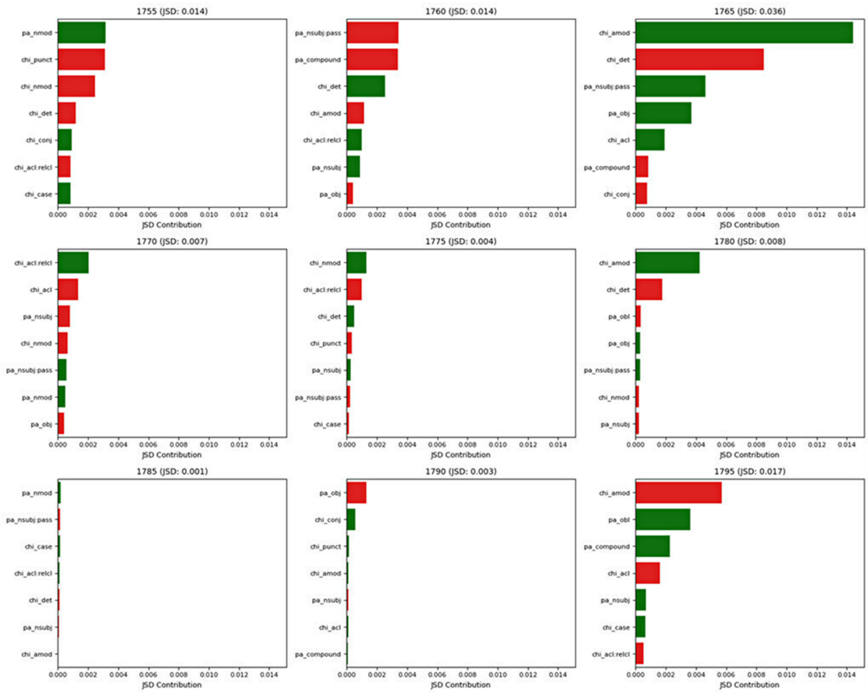


Figure 6 shows the Jensen-Shannon Divergence (JSD) values across successive five-year time bins from 1750 to 1800, measuring the distributional change in syntactic slots associated with *air*. The most noticeable pattern is a sharp increase in divergence around the 1760–1765 period, peaking just above 0.035. This suggests a substantial shift in the way *air* was used syntactically and coincides with the slot-filler adjective modifier as shown in the analysis of the diachronic distribution, reflecting the emergence of new descriptive terminology (e.g., *fixed air*, *dephlogisticated air*). Following this peak, JSD drops steeply from 1765 to 1770 and continues to decline to its lowest value around 1785. This period can be interpreted as a phase of relative stabilization, in how *air* was syntactically described, showing a phase of terminological consolidation as certain descriptions gained acceptance in the community. After 1785, divergence gradually increases again, with a marked increase between 1790 and 1795, likely corresponding to the transition toward oxygen-centered theory, as Lavoisier's ideas gained traction. Together, these fluctuations trace the linguistic signature of conceptual instability and eventual transformation.

However, note that the absolute value of the JSD was extremely small, which means that generally, the distributions were very stable. When visualizing the contributions of different slots, the picture becomes clearer. All the big increases in the JDS pattern were mostly contributed by the *amod* slot (Figure 7).

During the 1750s and early 1760s, *amod* contributed insignificantly. However, it contributed significantly to the big divergence increment in the late 1760s. For the next 30 years, the contribution became minimal again, except for the early 1780s. And suddenly, in another rise in divergence at the end of the period, its contribution rose again.

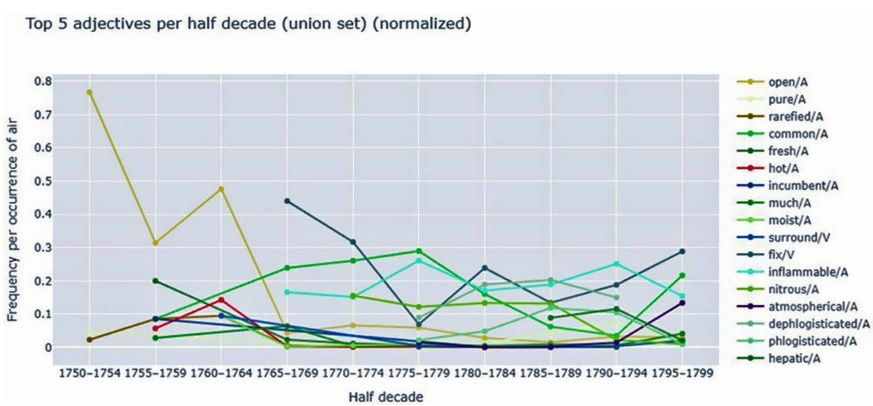
Figure 7: Contributions of different slots in the JSD



4.2.3 Assessing shifts in adjectival slot-filler distributions

We applied the same workflow with the adjective slot. First, we visualized the shift in relative frequencies of the adjectives and found that the discovery of oxygen changed the way scientists described *air*, causing great fluctuations of *technical* adjectives (e.g., *phlogisticated*, *fixed*) throughout the period (Figure 9). The frequencies of other *non-technical* adjectives (e.g., *fresh*, *open*) decreased and stayed low most of the time (Figure 8).

Figure 8: Relative frequency shifts of the adjectives filling the amod slot



Timewise, we can see some similarities in terms of the small periods with that of the above figures.

After computing the JSD score for the distributions of the adjectives, we found that it matches perfectly with that of the slots' JSD (Figure 10). The consistent pattern is: stability in the first decade, then a big rise followed by a stable period and an increase at the end.

Figure 9: The frequency shift of technical adjectives

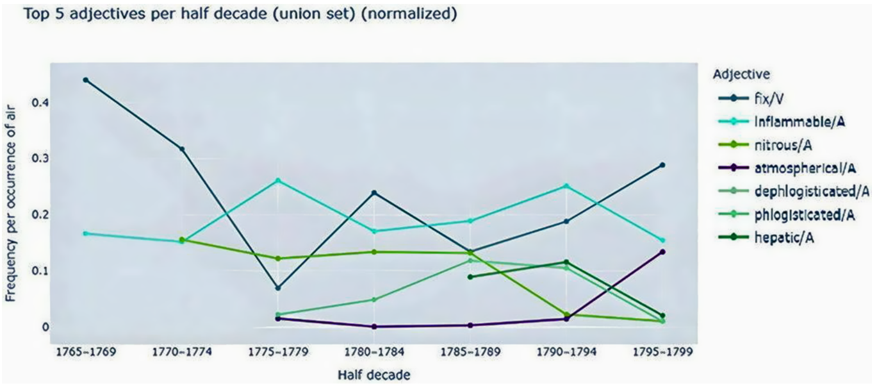


Figure 10: JSD of the adjectives

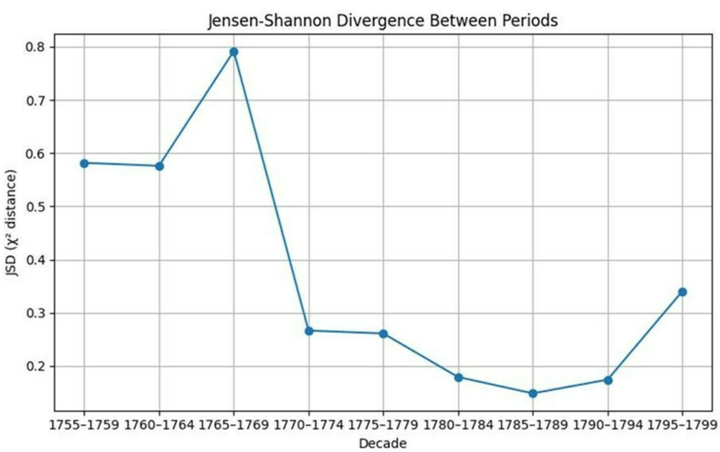
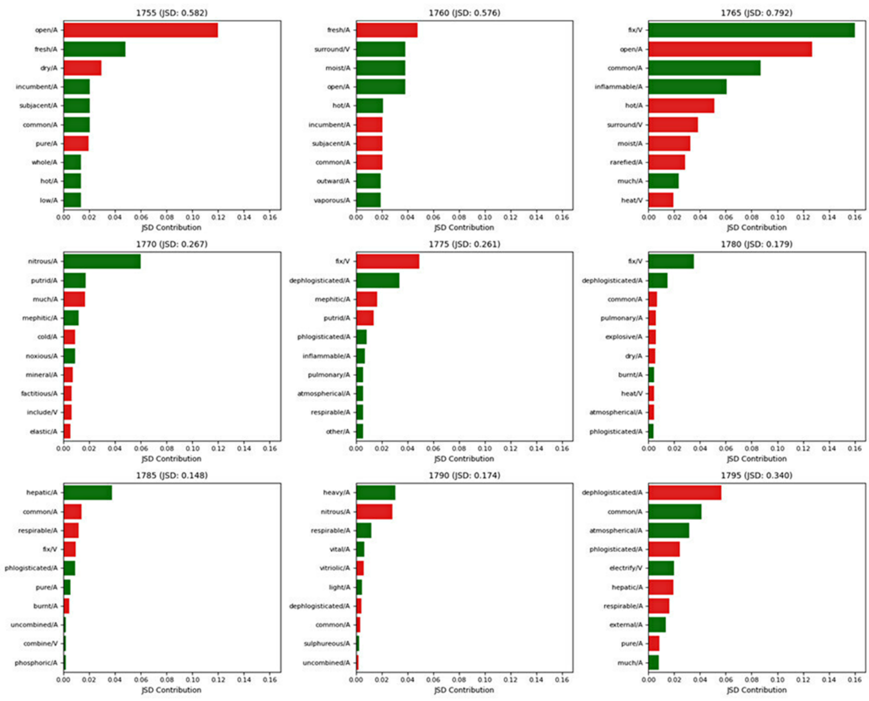


Figure 11 presents the JSD contributions of different adjectives modifying the noun *air* across 5-year intervals from 1755 to 1799. Each subplot represents a time bin and shows which adjectives (slot-fillers of the amod slot) contributed most to the distributional shift in that period. This fine-grained analysis allows us to trace how conceptual associations with *air* changed during the Chemical Revolution. The results converge with and help to quantify the linguistic aspect of the established narrative findings from the history and philosophy of science (Chang, 2011; Kuhn, 2012; McEvoy, 1978;

Stewart, 2012). Early adjectival modifiers reflect physical or environmental properties, while mid-century modifiers indicate taxonomization of different air types (e.g., *fixed*, *mephitic*), consistent with the exploratory period of phlogiston theory. In the 1770s and 1780s, the divergence score decreased considerably, indicating stabilisation around main competing descriptors (e.g., *(de-)phlogisticated*, *fixed*) at the height of debate. By the 1790s, phlogiston-related terms disappeared while compositional and functional terms dominated, such as *respirable* and *uncombined*, causing the divergence to rise again. This aligns with the defeat of phlogiston theory and the rise of oxygen theory. These changing slot-filler patterns offer linguistic evidence for the conceptual restructuring that defined the Chemical Revolution.

Figure 11: Contributions of different adjectives to the JSD



4.2.4 Exploratory comparison with LLM-based interpretations

Based on the findings, we explored whether commercial LLMs (ChatGPT and Grok), trained on large amounts of secondary scientific and historical material, would be able to provide similar results in terms of a comparable high-level narrative. To this end, we conducted a brief, illustrative comparison by prompting ChatGPT and Grok in a fresh session to avoid contamination from prior interactions. The convergence between the LLMs' responses and our finding lies in the identification of three broad periods characterized by epistemic struggle rather than abrupt replacement, as well as the prominence of key terms such as *(de)phlogisticated*, *common*, *fixed*. However, their responses only track macro-level conceptual change and largely reflect established philosophical and historical narratives, whereas our analysis provides a fine-grained, quantitative account grounded in tractable and transparent distributional and syntactic evidence. In this sense, LLMs appear well suited for exploratory purposes, but they do not (yet) replace the need for interpretable analyses when the goal is to trace how conceptual change is realized through language and to relate quantitative findings to a known and controllable data source.

4.3 Toward graph-based modeling of conceptual changes

The present study provides a first step towards modeling the multi-layered nature of conceptual changes in an interpretable, data-driven fashion. By combining topic modeling and syntactic slot-filler analysis across time, we have extracted core elements that can be operationalized as nodes and relationships in dynamic knowledge graphs. We envision these graphs to be able to represent the trajectories, tensions, and reconfigurations that constitute conceptual change in science as shown in recent studies (Kaye et al., 2024).

At the document level, our topic modeling reveals a sequential progression across three topics: *water* (1760s), *air* (1780s), and *acid* (1800s). These topics correspond to key stages in the theoretical shift from phlogiston to oxygen theory. In a graph-based framework, each of these topics can be modeled as a node anchored in time. The edges between them can encode their pairwise similarity using Jensen-Shannon Divergence (JSD) scores calculated from topic distributions. For instance, the edge between *water* (1760s) and *air* (1780s) would be weighted by their JSD, capturing the degree of topical shift. The same applies to the edge between *air* and *acid* (1800s), representing the continued drift in conceptual content (i.e. the meaning of *air* moving conceptually from *water* to *acid*). More importantly, the *air* node is not just one point on a timeline. It functions as a transitional hub. Based on our findings, *air* appears linguistically and conceptually between the decline of phlogiston (*water*) and the rise of oxygen (*acid*), both temporally and in terms of content similarity. The graph should reflect this mediating function explicitly: *air* is connected to both *water* and *acid*, and these connections are directional and weighted, signifying that *air* inherited semantic traits from one and contributed to the emergence of the other.

From the word-level analysis using SynFlow, we derive additional relational data that can be layered into the graph. For instance, the slot-filler analysis revealed that modifiers of *air* changed significantly over time. In the earlier period, *air* was often preceded by spatial or environmental adjectives (e.g., *open*, *fresh*), while in the later period, modifiers became chemically loaded (e.g., *vital*, *fixed*, *dephlogisticated*). This shift in the amod (adjective modifier) slot can be modeled as a series of edges connecting the *air* node to adjective nodes representing its contextually salient attributes. These attributes themselves can evolve and be clustered, representing the change in the conceptual schema surrounding *air*. Furthermore, we observed a decline in the relative frequency of the determiner slot (det) for *air*, suggesting a shift from referential usage (*the air*) to more abstract, substance-like usage without the determiner. In graph terms, this could be represented as a change in node type or ontological status, i.e. a transition from a generic, external referent (i.e. *the air*) to a theoretically posited scientific entity. This supports modeling *air* not just in semantic or distributional terms but also as a concept that changes category within the scientific ontology, something graph-based ontologies can represent explicitly.

Entropy values from the topic analysis add another dimension. High entropy in a topic suggests conceptual plurality or competition, while declining entropy reflects conceptual consolidation. Thus, entropy values can be used as node-level attributes or as weights modulating the confidence or centrality of a node at a given time. For example, the high entropy of the *air* topic during the 1780s may indicate a state of conceptual fluidity, just before *acid* (oxygen theory) became dominant in the 1800s.

All of these elements (topic trajectories, JSD values, syntactic environments, slot-filler shifts, determiner frequency changes, and entropy dynamics) can be encoded as graph elements. Topics such as *air*, *water*, and *acid*, as well as key lexical terms like *air*, can serve as nodes. These nodes may be enriched with associated properties derived from modifier terms extracted via syntactic analysis, such as *vital*, *fixed*, or *dephlogisticated*, which reflect the evolving semantic framing of these concepts. Additionally, rhetorical functions such as *observation* or *substance* identified in topic contexts could themselves be modeled as nodes, enabling the tracking of shifts in conceptual roles over time. Connections between these elements are encoded as edges. Topic-to-topic edges, for example, can be weighted using the JS distance to quantify thematic divergence across time. Term-to-term edges may be derived from co-modification patterns (e.g., when *air* and *gas* share similar adjectives), while term-to-property edges arise from slot-filler patterns (e.g., *air* modified by *vital* in the amod slot). Edges can also capture temporal progression, allowing chronological anchoring of conceptual developments. Each node can carry specific attributes that reflect the linguistic and conceptual state of the entity it represents. Entropy scores can capture topic diversity at a given time; relative frequency can indicate prominence or decline; grammatical role distributions (e.g., the presence or absence of determiners) can reflect shifts in ontological framing. The types of edges themselves can be semantically rich: they may represent semantic similarity (as computed from word embeddings or co-occurrence), rhetorical continuity (when one topic or term extends the discourse function of another), syntactic modification (capturing grammatical patterns), or ontological transitions (e.g., when *air* shifts from being treated as an environmental referent to a chemical entity). This formalization provides the structural backbone for graph-based modeling of conceptual shifts, allowing us to integrate linguistic observations into a dynamic and interpretable network of conceptual change. This graph could be dynamically constructed over time slices and queried to trace how concepts evolve, converge, or disappear.

Ultimately, the analyses presented here serve as a scaffolding: they identify the elements and dynamics that need to be encoded in future graph-based models. They also offer quantitative backing for where conceptual transitions occur and how they manifest in language. By translating these observations into graph structures, we aim to model conceptual shifts not as isolated changes but as dynamic, interconnected processes of semantic and epistemic evolution.

This view is similar to Vogl et al.'s (2026) where LLMs are described "*as a compass rather than a quick fix*". Although our tasks have different objects of study, the overarching goal of detecting shifts in discourse is the same. Their approach used topic modeling and LLMs combined to generate interpretable labels for a knowledge graph, and despite the challenges of characterizing the driving forces of the narratives or the participants, we are excited to explore akin implementations in future works.

5. Conclusion

This paper examined how conceptual change during the late phase of the Chemical Revolution is reflected in scientific writing by combining document-level topic modeling and word-level syntactic slot-filler analysis with information-theoretic measures to interpret change through linguistic insight. The results show a gradual reorganization with phases of thematic competition and consolidation at both levels. In particular, we observed the transition from phlogiston to oxygen theory, with topic shifts (*air* to *acid*) accompanied by higher entropy and slot-fillers with rising divergence during that period.

Beyond the historical case study, the contribution of this work lies in showing the value of interpretable, distribution-based methods for linguistically HPSS-oriented questions such as conceptual changes. Topic entropy and divergence make explicit how thematic diversity and stabilization unfold over time, while slot-filler patterns allow conceptual change to be traced through observable grammatical configurations. Compared to LLM-based approaches, which recover high-level narratives but conflate lexical, semantic, and grammatical information and do not offer direct access to the linguistic configurations through which change is enacted, our approach provides a transparent link between quantitative signals and the linguistic realization of epistemic change in a clearly defined primary corpus.

By translating these interpretable linguistic features into nodes, edges, and attributes, future work can integrate the strengths of distributional analysis with more expressive representational frameworks, potentially complementary with LLM-based methods. In this sense, the analyses presented here are not an endpoint, but a scaffolding for modeling conceptual change as a dynamic, multi-layered process grounded in observable language use.[4]

Acknowledgements

Funded by the European Union. Views and opinions ex-pressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held re-sponsible for them.

4 This chapter was written with support from large language models (LLMs). All model-generated text was reviewed and, where necessary, rewritten by the authors, who remain fully responsible for the final version. For details on the use of LLMs in this volume, see the statement in the volume's introduction.

References

Bengio Y, Ducharme R, Vincent P, et al. (2003) A neural probabilistic language model. J. Mach. Learn. Res. 3(null): 1137–1155.

Bensaude-Vincent B and Abbri F (1995) Lavoisier in European Context: Negotiating a New Language for Chemistry. Science History Publications/USA, Canton, Massachusetts.

Bizzoni Y, Degaetano-Ortlieb S, Menzel K, et al. (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 (eds S Degaetano-Ortlieb, A Kazantseva, N Reiter, et al.), Punta Cana, Dominican Republic (online), November 2021, pp. 120–127. Association for Computational Linguistics. Available at: https://aclanthology.org/2021.latechclfl-1.14

Blei DM, Ng AY and Jordan MI (2003) Latent dirichlet allocation. J. Mach. Learn. Res. 3(null): 993–1022.

Cassotti P, Periti F, De Pascale S, et al. (2024) Computational modeling of semantic change. In: Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts (eds M Mesgar and S Loáiciga), St. Julian's, Malta, March 2024, pp. 1–8. Association for Computational Linguistics. Available at: https://aclanthology.org/2024.eacl-tutorials.1/

Chang H (2011) The Persistence of Epistemic Objects Through Scientific Change. Erkenntnis 75(3): 413–429.

Degaetano-Ortlieb S and Teich E (2018) Using relative entropy for detection and analysis of periods of diachronic linguistic change. In: Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (eds B Alex, S Degaetano-Ortlieb, A Feldman, et al.), Santa Fe, New Mexico, August 2018, pp. 22–33. Association for Computational Linguistics. Available at: https://aclanthology.org/W18-4503

Degaetano-Ortlieb S and Teich E (2019) Toward an optimal code for communication: The case of scientific English. Corpus Linguistics and Linguistic Theory 18(1): 175–207.

Fankhauser P, Knappen J and Teich E (2016) Topical Diversification Over Time In The Royal Society Corpus. In: ADHO 2016 – Kraków, 2016. Available at: https://dh-abstracts.library.cmu.edu/works/2457

Firth JR (1957) A Synopsis of Linguistic Theory, 1930–1955. In: Studies in Linguistic Analysis. Oxford: Basil Blackwell. Available at: https://cs.brown.edu/courses/csci2952d/readings/lecture1-firth.pdf

Fischer S, Knappen J, Menzel K, et al. (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 (eds N Calzolari, F Béchet, P Blache, et al.), Marseille, France, May 2020, pp. 794–802. European Language Resources Association. Available at: https://aclanthology.org/2020.lrec-1.99/

Geeraerts D (1997) Diachronic Prototype Semantics: A Contribution Historical Lexicology. Oxford University Press. Available at: https://doi.org/10.1093/oso/9780198236528.001.0001

Geeraerts D (2009) Theories of Lexical Semantics. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780198700302.001.0001

Geeraerts D, Grondelaers S and Bakema P (2012) The Structure of Lexical Variation: Meaning, Naming, and Context. De Gruyter Mouton. Available at: https://www.degruyterbrill.com/document/doi/10.1515/9783110873061/html?lang=en

Griffiths TL and Steyvers M (2004) Finding scientific topics. Proceedings of the National Academy of Sciences 101(suppl_1). Proceedings of the National Academy of Sciences: 5228–5235.

Hall D, Jurafsky D and Manning CD (2008) Studying the history of ideas using topic models. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing – EMNLP '08, Honolulu, Hawaii, 2008, p. 363. Association for Computational Linguistics. Available at: http://portal.acm.org/citation.cfm?doid=1613715.1613763

Holmes FL (1994) What was the chemical revolution about. In: Abstracts of Papers of the American Chemical Society, USA, 1994, pp. 13–HIST. American Chemical Society.

Kaye A, Schlattmann R, Vogl M, et al. (2024) Socio-Epistemic Networks: A Framework for History of Knowledge. In: Lerg CA, Östling J, Weiß J, et al. (eds) History of Intellectual Culture 3/2024: Experimental Spaces: Knowledge Production and Its Environments in the Long Nineteenth Century. De Gruyter Oldenbourg, pp. 253–274. Available at: https://www.degruyterbrill.com/document/doi/10.1515/9783111291383-012/html

Kermes H, Degaetano-Ortlieb S, Khamis A, et al. (2016) The Royal Society Corpus: From Uncharted Data to Corpus. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (eds N Calzolari, K Choukri, T Declerck, et al.), Portorož, Slovenia, May 2016, pp. 1928–1931. European Language Resources Association (ELRA). Available at: https://aclanthology.org/L16-1305

Kuhn TS (1962) The Structure of Scientific Revolutions. University of Chicago Press: Chicago.

Kuhn TS (2012) The Structure of Scientific Revolutions. 50th Anniversary Edition. University of Chicago Press: Chicago.

Lareau F and Malaterre C (2026) When less is more? Topic model performance across title, abstract, and full-text corpora. In: Simons A, Wüthrich A, Zichert M, et al. (eds) *Understanding Science with Large Language Models? Potentials for the History, Philosophy, and Sociology of Science*. Bielefeld: transcript, part-4.

Luhmann J and Burghardt M (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.

McEvoy JG (1978) Joseph Priestley, 'Aerial philosopher': Metaphysics and methodology in Priestley's chemical thought, from 1772 to 1781, part I. Ambix 25(1): 1–55.

Menzel K, Knappen J, and Teich E (2021) Generating linguistically relevant metadata for the Royal Society Corpus. Research in Corpus Linguistics, 9(1). Available at: https://doi.org/10.22028/D291-38086

Mimno D, Wallach HM, Talley E, et al. (2011) Optimizing semantic coherence in topic models. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, USA, 27 July 2011, pp. 262–272. EMNLP '11. Association for Computational Linguistics.

Mohr JW and Bogdanov P (2013) Introduction—Topic models: What they are and why they matter. Poetics 41(6). Topic Models and the Cultural Sciences: 545–569.

Periti F and Montanelli S (2024) Lexical Semantic Change through Large Language Models: a Survey. ACM Comput. Surv. 56(11): 282:1-282:38.

Periti F and Tahmasebi N (2024) A Systematic Comparison of Contextualized Word Embeddings for Lexical Semantic Change. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (eds K Duh, H Gomez, and S Bethard), Mexico City, Mexico, June 2024, pp. 4262–4282. Association for Computational Linguistics. Available at: https://aclanthology.org/2024.naacl-long.240/

Periti F, Ferrara A, Montanelli S, et al. (2022) What is Done is Done: an Incremental Approach to Semantic Shift Detection. In: Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change (eds N Tahmasebi, S Montariol, A Kutuzov, et al.), Dublin, Ireland, May 2022, pp. 33–43. Association for Computational Linguistics. Available at: https://aclanthology.org/2022.lchange-1.4/

Phan-Tất B (2025) phantatbach/SynFlow: 20250739. Zenodo. Available at: https://zenodo.org/records/16610427 (accessed 21 October 2025).

Qi P, Zhang Yuhao, Zhang Yuhui, et al. (2020) Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Online, 2020, pp. 101–108. Association for Computational Linguistics. Available at: https://www.aclweb.org/anthology/2020.acl-demos.14

Schmidt BM (2012) Words Alone: Dismantling Topic Models in the Humanities. Journal of Digital Humanities 2(1): 49–65.

Shannon CE (1948) A Mathematical Theory of Communication. Bell System Technical Journal 27(3): 379–423.

Sievert C and Shirley K (2014) LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the workshop on interactive language learning, visualization, and interfaces, 2014, pp. 63–70.

Singh A, D'Arcy M, Cohan A, et al. (2023) SciRepEval: A Multi-Format Benchmark for Scientific Document Representations. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (eds H Bouamor, J Pino, and K Bali), Singapore, December 2023, pp. 5548–5566. Association for Computational Linguistics. Available at: https://aclanthology.org/2023.emnlp-main.338/

Stewart J (2012) The Reality of Phlogiston in Great Britain. Hyle 18(2): 175–194.

Tahmasebi N, Borin L, Jatowt A, et al. (2021) Computational Approaches to Semantic Change. Language Science Press. Available at: https://langsci-press.org/catalog/book/303

Thagard P (1990) The Conceptual Structure of the Chemical Revolution. Philosophy of Science 57(2). [The University of Chicago Press, Philosophy of Science Association]: 183–209.

Vogl M, von Schwerin A and Kirschke S (2026) Large language models in interdisciplinary research settings. A reflection. In: Simons A, Wüthrich A, Zichert M, et al. (eds) *Understanding Science with Large Language Models? Potentials for the History, Philosophy, and Sociology of Science*. Bielefeld: transcript, part-4.