



Linguistic Variation and Change in 250 Years of English Scientific Writing: A Data-Driven Approach

Yuri Bizzoni^{1*}, Stefania Degaetano-Ortlieb¹, Peter Fankhauser² and Elke Teich¹

¹ Language Science and Technology, Saarland University, Saarbrücken, Germany, ² Digital Linguistics, Institut für Deutsche Sprache, Mannheim, Germany

We trace the evolution of Scientific English through the Late Modern period to modern time on the basis of a comprehensive corpus composed of the Transactions and Proceedings of the Royal Society of London, the first and longest-running English scientific journal established in 1665. Specifically, we explore the linguistic imprints of specialization and diversification in the science domain which accumulate in the formation of “scientific language” and field-specific sublanguages/registers (chemistry, biology etc.). We pursue an exploratory, data-driven approach using state-of-the-art computational language models and combine them with selected information-theoretic measures (entropy, relative entropy) for comparing models along relevant dimensions of variation (time, register). Focusing on selected linguistic variables (lexis, grammar), we show how we deploy computational language models for capturing linguistic variation and change and discuss benefits and limitations.

Keywords: linguistic change, diachronic variation in language use, register variation, evolution of Scientific English, computational language models

OPEN ACCESS

Edited by:

David Jurgens,
University of Michigan, United States

Reviewed by:

Joshua Waxman,
Yeshiva University, United States
Yao-Ying Lai,
National Chengchi University, Taiwan

*Correspondence:

Yuri Bizzoni
yuri.bizzoni@uni-saarland.de

Specialty section:

This article was submitted to
Language and Computation,
a section of the journal
Frontiers in Artificial Intelligence

Received: 03 February 2020

Accepted: 07 August 2020

Published: 16 September 2020

Citation:

Bizzoni Y, Degaetano-Ortlieb S,
Fankhauser P and Teich E (2020)
Linguistic Variation and Change in 250
Years of English Scientific Writing: A
Data-Driven Approach.
Front. Artif. Intell. 3:73.
doi: 10.3389/frai.2020.00073

1. INTRODUCTION

The language of science is a socio-culturally firmly established domain of discourse that emerged in the Early Modern period (ca. 1500–1700) and fully developed in the Late Modern period (ca. 1700–1900). While considered fairly stable linguistically (cf. Görlach, 2001; Leech et al., 2009), the Late Modern period is a very prolific time when it comes to the formation of text types, with many of the registers we know today developing during that period—including the language of science (see Görlach, 2004 for a diachronic overview).

Socio-culturally, register diversification is connected to the growing complexity of modern societies, labor becoming increasingly divided with more different and increasingly specialized activities across all societal sectors¹. Also, driven by science as well as early industry, standardization (e.g., agreements on weights and measures) and routinization of procedures become important issues. At the same time, enlightenment and the scientific and industrial revolutions support a general climate of openness and belief in technological advancement. In the domain of science, the eighteenth century is of course the epoch of encyclopedias² but also that of the scientific academies which promoted the scientific method and distributed scientific knowledge through

¹An example in point are production and experimentation, which used to be carried out hand in hand in the workshops of alchemists and apothecaries but were separated later on, also physically, with experimentation becoming a scientific activity carried out in dedicated laboratories (Burke, 2004; Schmidgen, 2011).

²For example, the publication of the famous *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers* (1751–1765).

their publications. The two oldest scientific journals are the French *Journal des Sçavans* and the *Philosophical Transactions of the Royal Society of London*. At the beginning of publication (both started in 1665), the journals were no more than pamphlets and included articles written in the form of letters to the editor and reviews of scientific works (Gleick, 2010). Professionalization set in around the mid eighteenth century, as witnessed by the introduction of a reviewing process in the Royal Society (Moxham and Fyfe, 2018; Fyfe et al., 2019).

While there is a fair stock of knowledge on the development of scientific language from socio-cultural and historical-pragmatic perspectives (see section 2), it is less obvious what are the underlying, more general principles of linguistic adaptation to new needs of expression in an increasingly diversified and specialized setting such as science. This provides the motivation for the present research. Using a comprehensive diachronic corpus of English scientific writing composed of the *Philosophical Transactions* and *Proceedings of the Royal Society of London* [henceforth: Royal Society Corpus (RSC); Kermes et al., 2016; Fischer et al., 2020], we trace the evolution of Scientific English looking for systematic linguistic reflexes of specialization and diversification, yielding a distinctive “scientific style” and forming diverse sublanguages (sublanguage of chemistry, physics, biology etc.). In terms of theory, our work is specifically rooted in register linguistics (Halliday, 1985b; Biber, 1988) and more broadly in theories of language use, variation and change that acknowledge the interplay of social, cognitive and formal factors (e.g., Bybee, 2007; Kirby et al., 2015; Aitchison, 2017; Hundt et al., 2017). While we zoom in on the language of science, we are ultimately driven by the more general questions about language change: What changes and how? What drives change? How does change proceed? What are the effects of change? Thus, we aim at general insights about the dynamics of language use, variation and change.

In a similar vein, the methodology we present can be applied to other domains and related analysis tasks as well as other languages. Overall, we pursue an exploratory, data-driven approach using state-of-the-art computational language models (ngram models, topic models, word embeddings) combined with selected information-theoretic measures (entropy, relative entropy) to compare models/corpora along relevant dimensions of variation (here: time and register) and to interpret the results with regard to effects on language system and use. Since the computational models we use are word-based, words act as the anchor unit of analysis. However, style is primarily indicated by lexico-grammatical usage, so we investigate both the lexical and the grammatical side of words. While we consider lexis and grammar as intricately interwoven, in line with various theories of grammar (Halliday, 1985a; Hunston and Francis, 2000; Goldberg, 2006), for expository purposes, we here consider the lexico-semantic and the lexico-grammatical contributions to change separately.

The remainder of the paper is organized as follows. We start with an overview of previous work in corpus and computational linguistics in modeling diachronic change with special regard to register and style (section 2). In section 3 we introduce our data set (section 3.1) and elaborate on the methods employed (section

3.3). Section 4 shows analyses of diachronic trends at the levels of lexis and grammar (section 4.1), the development of topics over time (section 4.2) and paradigmatic effects of changing language use (section 4.3). Finally, we summarize our main results and briefly assess benefits and shortcomings of the different kinds of models and measures applied to the analysis of linguistic variation and change (section 5).

2. RELATED WORK

The present work is placed in the area of language variation and change with special regard of social and register variation and computational models of variation and change (for overviews see Aragamon, 2019 for computational register studies and Nguyen et al., 2016 for computational socio-linguistics).

Regarding the language of science, there is an abundance of linguistic-descriptive work, including diachronic aspects, providing many valuable insights (e.g., Halliday, 1988; Halliday and Martin, 1993; Atkinson, 1999; Banks, 2008; Biber and Gray, 2011, 2016). However, most of the existing work is either based on text samples or starts from predefined linguistic features. Further, there are numerous studies on selected scientific domains, such as medicine or astronomy, e.g., Nevalainen (2006), Moskowich and Crespo (2012) and Taavitsainen and Hiltunen (2019), which work on the basis of fairly small corpora containing hand-selected and often manually annotated material. Typically, such studies are driven from a historical socio-linguistic or pragmatic perspective and focus on selected linguistic phenomena, e.g., forms of address (Taavitsainen and Jucker, 2003). For overviews on recent trends in historical pragmatics/socio-linguistics (see Jucker and Taavitsainen, 2013; Säily et al., 2017). Studies on specific domains, registers or text types provide valuable resources and insights into the socio-historical conditions of language use. Here, we build upon these insights, adding to it the perspective of general mechanisms of variation and change.

More recently, the diachronic perspective has attracted increasing attention in computational linguistics and related fields. Generally, diachronic analysis requires a methodology for comparison of linguistic productions along the time line. Such comparisons may range over whole epochs (e.g., systemic changes from early Modern English to Late Modern English), or involve short ranges (e.g., the issues of 1 year of *The New York Times* to detect topical trends). Applying computational language models to diachronic analysis requires a computationally valid method of comparison of language use along the time line, i.e., one that captures linguistic change if it occurs.

Different kinds of language models are suitable for this task and three major strands can be identified. First, a number of authors from fields as diverse as literary studies, history and linguistics have used simple ngram models to find trends in diachronic data using relative entropy (Kullback-Leibler Divergence, Jensen-Shannon Divergence) as a measure of comparison. For instance, Juola (2003) used Kullback-Leibler Divergence (short: KLD) to measure rate of linguistic change in

30 years of National Geographic Magazine. In more recent, large-scale analyses on the Google Ngram Corpus (Bochkarev et al., 2014; Kim et al., 2014) analyze change in frequency distributions of words within and across languages. Specifically humanistic research questions are addressed by e.g., Hughes et al. (2012) who use relative entropy to measure stylistic influence in the evolution of literature; or Klingenstein et al. (2014) who analyze different speaking styles in criminal trials comparing violent with non-violent offenses; or Degaetano-Ortlieb and Teich (2018) applying KLD as dynamic slider over the time line of a diachronic corpus of scientific text.

Second, probabilistic topic models (Steyvers and Griffiths, 2007) have become a popular means to summarize and analyze the content of text corpora, including topic shifts over time. In linguistics and the digital humanities, topic models have been applied to various analytic goals including diachronic linguistic analysis (Blei and Lafferty, 2006; Hall et al., 2008; Yang et al., 2011; McFarland et al., 2013). Here again, a valid method of comparing model outputs along the time line has to be provided. In our work, we follow the approach proposed in Fankhauser et al. (2016) using entropy over topics as a measure to assess topical diversification over time.

Third, word embeddings have become a popular method for modeling linguistic change, with a focus on lexical semantic change (e.g., Hamilton et al., 2016; Dubossarsky et al., 2017, 2019; Fankhauser and Kupietz, 2017). Word embeddings are weakly neural models that capture usage patterns of words and are used in a variety of NLP tasks. While well-suited to capture the summative effects of change (groups of words or whole vocabularies, see e.g., Grieve et al., 2016), the primary focus lies on lexis³. Other linguistic levels, e.g., grammar (Degaetano-Ortlieb and Teich, 2016, 2018; Bizzoni et al., 2019a), collocations (Xu and Kemp, 2015; Garcia and García-Salido, 2019), or specific aspects of change, e.g., spread of change (Eisenstein et al., 2014), specialization (Bizzoni et al., 2019b) or life-cycles of language varieties (Danescu-Niculescu-Mizil et al., 2013), are only rarely considered. Once again, while word embeddings offer a specific model of language use, using them to capture diachronic change and to assess effects of change calls for adequate instruments for comparison along the time line. Here, we use the commonly applied measure of cosine distance for a general topological analysis of diachronic word embedding spaces; and we use entropy for closer inspection of specific word embeddings clusters to measure the more fine-grained paradigmatic effects of change.

In sum, in this paper we address some of the core challenges in modeling diachronic change by (a) looking at the *interplay* of different linguistic levels (here: lexis and grammar), (b) elaborating on the formation of style and register from a diachronic perspective, and (c) enhancing existing computational methods with explicit measures of linguistic change. Since

³For more comprehensive overviews on computational approaches to lexical semantic change see Tahmasebi et al. (2018) and on diachronic word embeddings see Kutuzov et al. (2018).

TABLE 1 | Size of RSC 6.0 by 50-year periods.

Time	# Tokens	# Texts
1665–1699	2,582,856	1,325
1700–1749	3,414,795	1,686
1750–1799	6,342,489	1,819
1800–1849	9,112,274	2,774
1850–1899	36,993,412	6,754
1900–1919	19,273,112	3,049

we are driven by the goal of explanation rather than high-accuracy prediction (as in NLP tasks), qualitative interpretation by humans is an integral step. Here, micro-analytic and visual support are doubly important if one wants to explore linguistic conditions and effects of change. To support this, good instruments for human inspection and analysis of data are crucial—see, for instance, Jurish (2018) and Kaiser et al. (2019) providing visualization tools for various aspects of diachronic change, partly with interactive function (Fankhauser et al., 2014; Fankhauser and Kupietz, 2017); or Hilpert and Perek (2015)'s application of motion charts to the analysis of meaning change. We developed a number of such visualization tools made available as web applications for inspection of the Royal Society Corpus (cf. section 3).

3. DATA AND METHODS

3.1. Data

The corpus used for the present analysis is the Royal Society Corpus 6.0 (Fischer et al., 2020). The full version is composed of the Philosophical Transactions and Proceedings of the Royal Society from 1665 to 1996. In total, it contains 295,895,749 tokens and 47,837 documents. Here, we use a version that is open-source under a creative commons license covering the period of 1665 to 1920. In terms of periods of English, this reflects the Late Modern period (1700–1900) plus a bit of material from the last decades of the Early Modern period (before 1700) as well as a number of documents from modern English. Altogether this open version contains 78,605,737 tokens and 17,520 documents.

Note that the RSC is not balanced, later periods containing substantially more material than earlier ones (see **Table 1**), which calls for caution regarding frequency effects. Other potentially interesting features of the corpus are that the number of different authors increases over time; so does the number of papers with more than one author.

The documents in the corpus are marked up with meta-data including author, year of publication, text type and time period (1-, 10-, 50-year periods). The corpus is tokenized, lemmatized, annotated with part-of-speech tags and normalized (keeping both normalized and original word forms) using standard tools (Schmid, 1995; Baron and Rayson, 2008). The corpus is made available under a Creative Commons license, downloadable and

accessible via a web concordance (CQPWeb; Hardie, 2012) as well as interactive visualization tools⁴.

3.2. Methods

There are two important a priori considerations regarding modeling linguistic change and variation. First, one of the key concepts in language variation is *use in context*. Apart from extra-linguistic, situational context (e.g., field, tenor, and mode; Quirk et al., 1985), intra-linguistic context directly impacts on linguistic choice, both syntagmatically (as e.g., in collocations) and paradigmatically (i.e., shared context of alternative expressions). Different computational models take into account different types of context and accordingly reveal different kinds of linguistic patterns. Topic models take into account the distribution of words in document context and are suitable to capture the field of discourse (see section 3.2.2 below). Plain ngram models take into account the immediately preceding words of a given word and can reveal syntagmatic usage patterns (see section 3.2.1 below). Word embeddings take into account left and right context (e.g., \pm five words) and allow clustering words together depending on similar, surrounding contexts; thus, they are suited for capturing linguistic paradigms (see section 3.2.3 below).

Second, diachronic linguistic analysis essentially consists of *comparison of corpora* representing language use at different time periods. Computational language models being representations of corpora, the core task consists in comparing model outputs and elicit significant differences between them. Common measures of comparing language models are perplexity and relative entropy, typically used for assessing the quality or fit of a model by estimating the difference between models in bits using a log base. Here, we use the asymmetric version of relative entropy, Kullback-Leibler Divergence, to assess differences between language models according to time. An intimately related measure is entropy. Entropy considers the richness and (un)evenness of a sample and is a common means to measure diversity, e.g., the lexical diversity of a language sample (Thoiron, 1986). Here, we use entropy as a measure of diversification at two levels, the level of topics (field of discourse) and the level of paradigmatic word clusters, where greater entropy over time is interpreted as a signal of linguistic diversification and lower entropy as a signal of consolidated language use. The most basic way of exploring change in a given data set is to test whether the entropy over a simple bag-of-words model changes or not. For diversification to hold, we would expect the entropy to rise over time in the RSC, also because of the increase in size of the more recent corpus parts as well as in number of authors. As will be seen, this is not the case, entropy at this level being fairly stable (section 4.2).

3.2.1. Ngram Based Models

To obtain a more fine-grained and linguistically informed overview of the overall diachronic tendencies in the RSC than possible with token ngrams, we consider lexical and grammatical usage separately using lemmas and part-of-speech (POS) sequences as modeling units. On this basis, models

of different time periods (e.g., decades) are compared with the asymmetric variant of relative entropy, Kullback-Leibler Divergence (KLD; Kullback and Leibler, 1951); cf. Equation (1) where A and B here denote different time periods.

$$D(A||B) = \sum_i p(\text{unit}_i|A) \log_2 \frac{p(\text{unit}_i|A)}{p(\text{unit}_i|B)} \quad (1)$$

KLD is a common measure for comparing probability distributions in terms of the number of additional bits needed for encoding when a non-optimal model is used. Applied to diachronic comparison, we obtain a reliable index of difference between two corpora A and B: the higher the amount of bits, the greater the diachronic difference. Also, we know which specific units/features contribute to the overall KLD score by their pointwise KLD. Thus, we can inspect particular points in time (e.g., by ranking features by pointwise KLD in 1 year) or time spans (e.g., by standard deviation across several years) to dynamically observe changes in a feature's contribution. This gives us two advantages over traditional corpus-based approaches: no predefined features are needed and results are more directly interpretable.

Apart from comparing predefined time periods with each other as is commonly done in diachronic corpus-linguistic studies (cf. Nevalainen and Traugott, 2012 for discussion), KLD can be used as a data-driven periodization technique (Degaetano-Ortlieb and Teich, 2018, 2019). KLD is dynamically pushed over the time line comparing past and future (or, as KLD is asymmetric, future vs. past). As we will show below, using KLD in this way allows detecting diachronic trends that are hard to see on a token level or with predefined, more coarse time periods. The granularity of diachronic comparison can be varied depending on the corpus and the analytic goal (year-, month-, day-based productions); again, no a priori assumptions have to be made regarding the concrete linguistic features involved in change other than selecting the linguistic level of comparison (e.g., lemmas, parts of speech). Hence, the method is generic and at the same time sensitive to the data.

3.2.2. Topic Models

To obtain a picture of the diachronic development in terms of field of discourse—a crucial component in register formation—we need to consider the usage of words in the context of whole documents. To this end, we use topic models. We follow the overall approach of applying topic models to diachronic corpora mapping topics to documents (Blei and Lafferty, 2006; Steyvers and Griffiths, 2007; Hall et al., 2008; Yang et al., 2011; McFarland et al., 2013). The principle idea is to model the generation of documents with a randomized two-stage process: For every word w_i in a document d select a topic z_k from the document-topic distribution $P(z_k|d)$ and then select the word from the topic-word distribution $P(w_i|z_k)$. Consequently, the document-word distribution is factored as: $P(w_i|d) = \sum_k P(w_i|z_k)P(z_k|d)$. This factorization effectively reduces the dimensionality of the model for documents, improving their interpretability: Whereas $P(w_i|d)$ requires one dimension for each distinct word (tens of thousands) per document, $P(z_k|d)$

⁴RSC 6.0 Open: <https://hdl.handle.net/21.11119/0000-0004-8E37-F>.

only requires one dimension for each topic (typically in the range of 20–100). To estimate the document-topic and topic-word distributions from the observable document-word distributions we use Gibbs-Sampling as implemented in MALLET⁵.

To investigate topical trends over time, we average the document-topic distributions for each year y :

$$P(z_k|y) = 1/n \sum_{d_j \in y} P(z_k|d_j) \quad (2)$$

where n is the number of documents per year.

For further interpretation, we cluster topics hierarchically on the basis of the distance⁶ between their topic-document distributions (Equation 3).

$$P(d|z) = P(z|d) / \sum_j P(z|d_j) \quad (3)$$

Topics that typically co-occur in documents have similar topic-document distributions, and thus will be placed close in the cluster tree.

To assess diachronic diversification in discourse field as a central part of register formation, we measure the entropy over topics (cf. Equation 4), and the mean entropy of topic-word distributions per time period.

$$H(P(\cdot|y)) = - \sum_k P(z_k|y) \log_2 P(z_k|y) \quad (4)$$

Note that all measures operate on relative frequencies per time period in order to control for the lack of balance in our data set (more recent periods contain considerably more data than earlier ones).

3.2.3. Word Embeddings

Word embeddings (WES) capture lexical paradigms, i.e., sets of words sharing similar syntagmatic contexts. Word embeddings build on the principle underlying distributional semantics that it is possible to capture important aspects of the semantics of words by modeling their context (Harris, 1954; Lenci, 2008).

Here, we apply WES diachronically to explore the overall development of word paradigms in our corpus with special regard to register/sublanguage formation as well as scientific style. Using the approach and tools provided by Fankhauser and Kupietz (2017) we compute WES with a structured skip-gram approach (Ling et al., 2015). This is a variant of the popular Word2Vec approach (Mikolov et al., 2013). Word2Vec is a way of maximizing the likelihood of a word given its context, by training a $d \times V$ matrix where V is the vocabulary and d an arbitrary number of dimensions.

The goal of the algorithm is to maximize

$$L = \frac{1}{T} \sum_{t \in T} \sum_{-c \leq j \leq c} \log p(w_{t+j}|w_t) \quad (5)$$

⁵<http://mallet.cs.umass.edu>

⁶We use Pearson distance, which consistently results in more intuitive hierarchies than Jensen-Shannon Divergence.

where T is a text and c is the number of left and right context words to be taken into consideration. In short, the model tries to learn the probability of a word given its context, $p(w_o|w_i)$. To this end, the model learns a set of weights that maximizes the probability of having a word in a given context. Such set of weights constitutes a word's embedding.

Usually, skip-gram considers a term's context as a bag-of-words. In Ling et al. (2015)'s variant, the order of the word context is also taken into consideration which is important to capture words with grammatical functions rather than lexical words only. For diachronic application, we calculate WES per time period (e.g., 1-/10-/50-year periods), where the first period is randomly initialized, and each subsequent period is initialized by the model for its preceding period. Thereby, WES are comparable across periods.

To perform analyses on our models, we then apply simple similarity measures commonly used in distributional semantics, where the similarity between two words is assessed by the cosine similarity of their vectors:

$$\text{sim}(w_1, w_2) = \cos(w_1, w_2) = \frac{w_1 w_2}{|w_1||w_2|} \quad (6)$$

where w_1 and w_2 are the vectors of the two words taken into consideration, and $|w|$ is a vector's norm. Alternatively, the semantic distance between words can be considered, which is the complement of their similarity:

$$\text{dist}(w_1, w_2) = 1 - \cos(w_1, w_2) \quad (7)$$

To detect the semantic tightness or level of clustering of a group of words (how semantically similar they are), one can thus compute the average cosine similarity between all the words in a group of words:

$$\text{sim}(V) = \frac{\sum_{w_a \in V} \sum_{w_b \in V} \cos(w_a, w_b)}{V^2} \quad (8)$$

where V (vocabulary) is the group of words taken into consideration. Reversely, it is possible to compute the average distance of a group of words from another group of words by iterating the sums on two different sets.

To detect semantic shifts over time, one of the simplest and most popular approaches is that of computing the change of the cosine similarity between a group of pre-defined words in a chronologically ordered set of WE spaces. As we will show, the WE space of the RSC as a whole expands over time. At the same time, it becomes more *fragmented* and specific clusters of words become more densely populated while others disappear. We base such observations on an analysis of the word embeddings' topology using cosine similarity as explained above as well as entropy. For example, since the period under investigation witnesses the systematization of several scientific disciplines, we are likely to observe a narrowing of the meaning of many individual words—mainly technical terms—which would push them further away from one another. Similarly, for specific WE clusters, we expect growth or decline, e.g., chemical terms explode in the late eighteenth century, pointing to the emergence

of the field of chemistry with the associated technical language, or many Latin words disappear. Such developments can be measured by the entropy $H(P(.|w))$ over a given cluster around word w , by estimating the conditional probability of words w_i in the close neighborhood of word w as follows:

$$P(w_i|w) = \text{sim}(w, w_i) * \text{freq}(w_i, w) / \left(\sum_k \text{sim}(w, w_k) * \text{freq}(w_k, w) \right) \quad (9)$$

where w_k ranges over all words (including w) with sufficient similarity (e.g., > 0.6) to w . The neighbors are weighted by their similarity to the given word, thus, a word with many near neighbors and rather uniform distribution has a large entropy, indicating a highly diversified semantic field.

4. ANALYSES

Our analyses are driven by two basic assumptions: register diversification (linguistic variation focused on field of discourse) and formation of “scientific style” (convergence on specific linguistic usages within the scientific domain). We carry out three kinds of analysis on the Royal Society Corpus showing these two major diachronic trends at the levels of lexis and grammar (section 4.1), development of topic over time (section 4.2) as well as paradigmatic effects (section 4.3).

4.1. Diachronic Trends in Lexis and Grammar

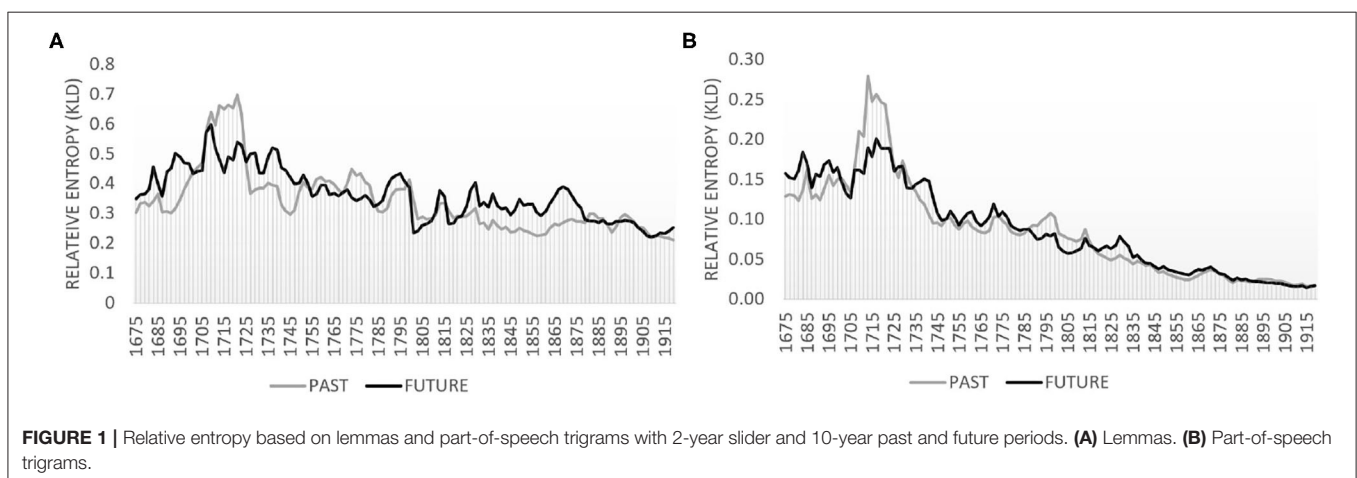
We trace the overall diachronic development in the RSC considering both lexical and grammatical levels. Lexis is captured by lemmas and grammar by sequences of three parts of speech (POS). Using the data-driven periodization technique described in section 3.2.1 based on KLD, we dynamically compare probability distributions of lemma unigrams and POS trigrams along the time line.

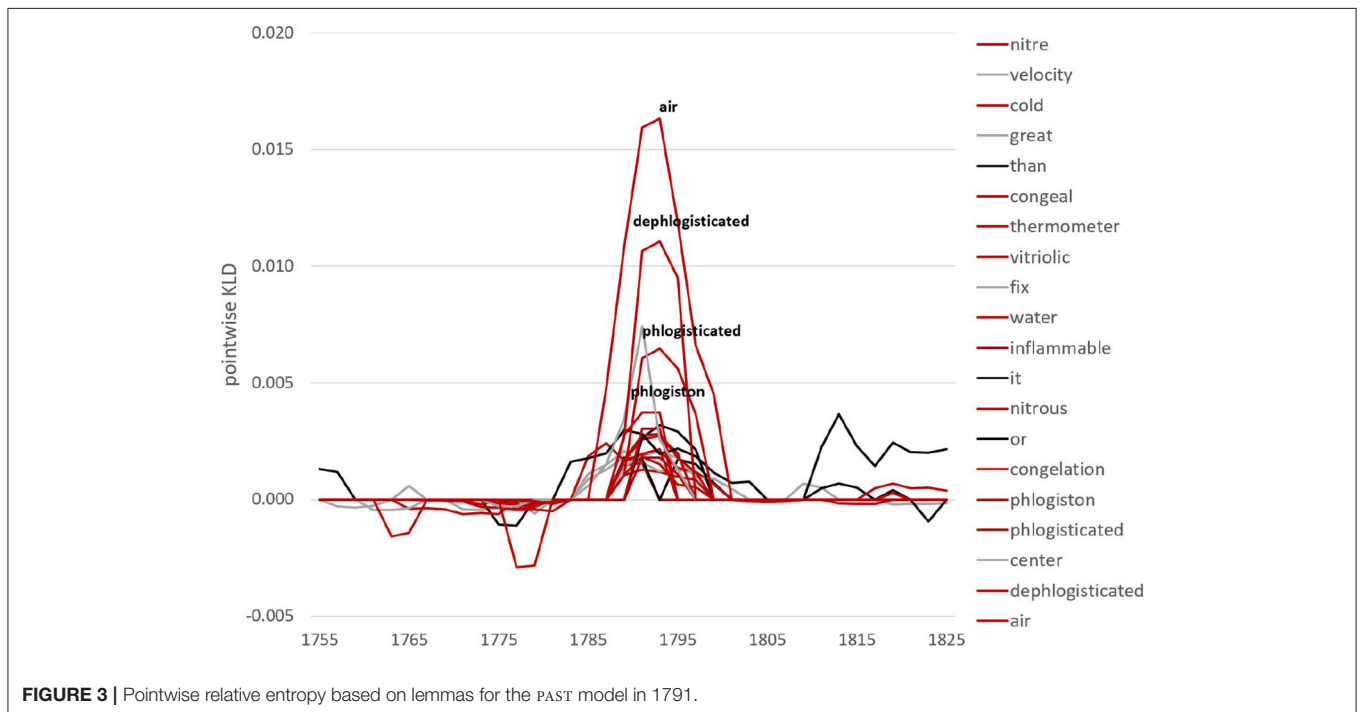
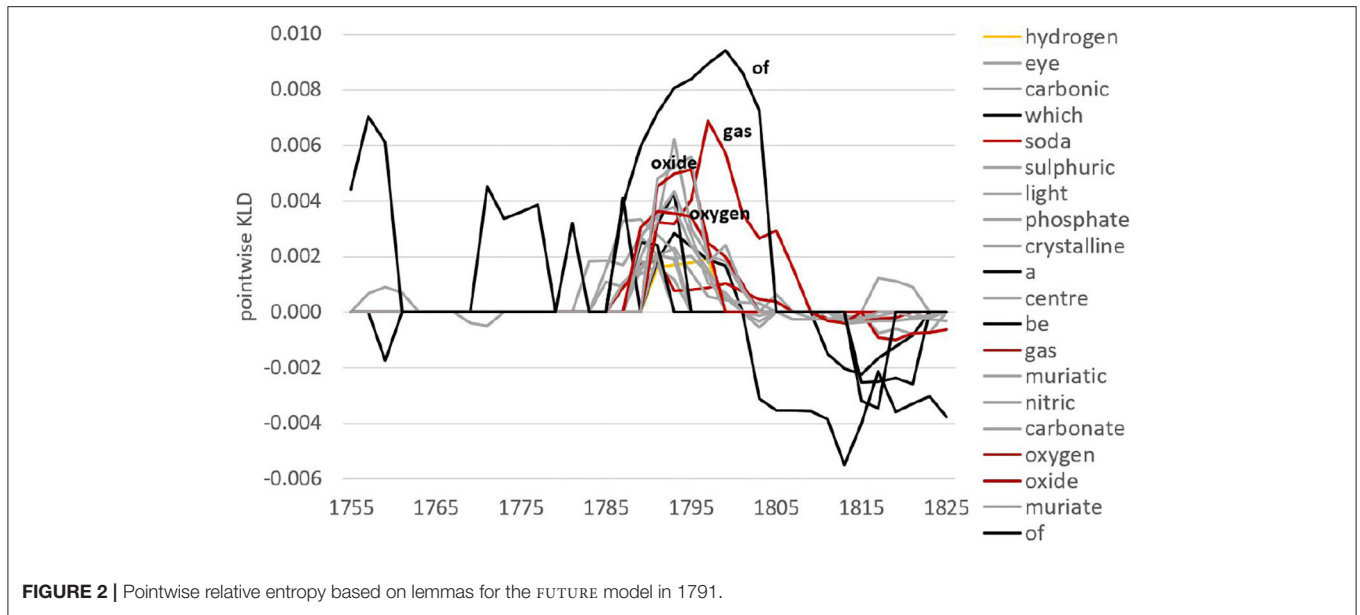
Figures 1A,B plot the temporal development for the lexical and the grammatical level, respectively. The black line visualizes relative entropy of the future modeled by the past, i.e., how well at a particular point in time the future can be modeled by a

model of the past (here: 10 year slices). The gray line visualizes the reverse, i.e., how well the past is modeled by the future (again on 10-year slices). Peaks in the black line indicate changes in the future which are not captured by a model of the past, such as new terminology. Peaks in the gray line indicate differences from the opposite perspective, i.e., the future not encompassing the past, e.g., obsolete terminology. Troughs for both lines indicate convergence of future and past. A fairly persistent, low-level relative entropy indicates a period of stable language use.

Comparing the two graphs in **Figure 1**, we observe a particularly strong decreasing tendency for the grammatical level (see **Figure 1B**) and a slightly declining tendency at the lexical level with fairly pronounced oscillations of peaks and troughs (**Figure 1A**). Basically, peaks indicate innovative language use, troughs indicate converging use, the future being less and less “surprised” by the past. Thus, while grammatical usage consolidates over time, the lexical level is more volatile as it reacts directly to the pressure of expressing newly emerging things or concepts in peoples’ (changing) domains of experience (here: new scientific discoveries). The downward trend at the grammatical level is a clear sign of convergence, possibly related to the formation of a scientific style; peaks at the lexical level signal innovative use and may indicate register diversification.

To investigate this in more detail, we look at specific lexical and grammatical developments. We use pointwise KLD (i.e., the contribution of individual features to overall KLD) to rank features. For example, there is a major increase in overall KLD around the 1790s at the lemma level. Considering features contributing to the highest peak in 1791 for the FUTURE model (black line), we see a whole range of words from the chemistry field around *oxygen* (see **Figure 2**). At the same time, we can inspect which features leave language use and contribute to an increase in KLD for the PAST model (i.e., features not well-captured by the future anymore). From **Figure 3**, we observe words related to *phlogiston* and experiments with air contributing to the formation of the oxygen theory of combustion (represented by Lavoisier, Priestley as well as Scheele). In fact, the oxygen theory replaced Becher and Stahl’s 100-years



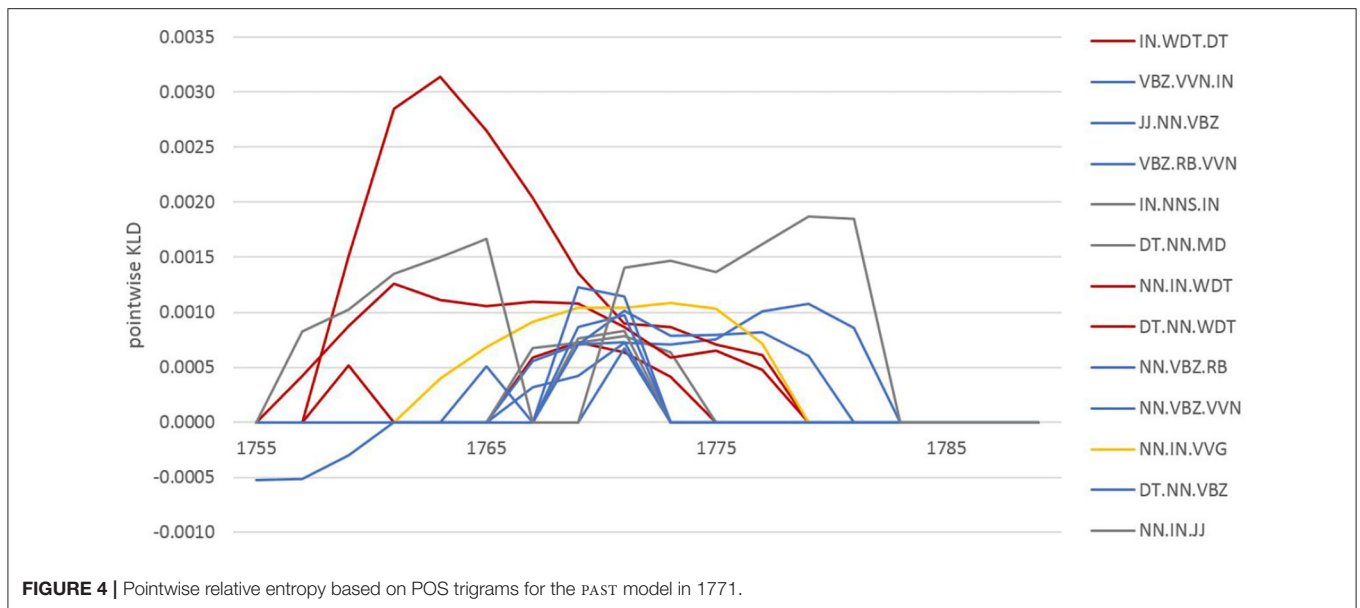


old phlogiston theory, marking a chemical revolution in the eighteenth century—it is this shift of scientific paradigm that we encounter here in the RSC.

At the grammatical level, after a fairly high KLD peak in the early 1700's, there is a step-wise, steady decrease with only local, smaller peaks. As an example of a typical development at the grammatical level consider the features involved in the 1771 peak (see **Figure 4**). These are passive voice and relational verb patterns (e.g., NOUN-BE-PARTICIPLE as in *air is separated*; blue), nominal patterns with prepositions [e.g., indicating

measurements such as the NOUN-PREPOSITION-ADJECTIVE as in *the quantity of common (air)*; gray], gerunds (e.g., NOUN-PREPOSITION-*ing*VERB, such as *method of making*; yellow), and relative clauses (e.g., DETERMINER-NOUN-RELATIVIZER, such as *the air which/that*; red). While the contribution of these patterns to the overall KLD is high in 1771, it becomes zero for all of them by 1785—a clear indication of consolidation in grammatical usage pointing to the development of a uniform scientific style.

Regarding the lexical level, to verify that the observed tendencies point to significant diversification effects, we need to



explore the systematic association of words with discourse fields. For this, we turn to topic models.

4.2. Diachronic Development of Discourse Fields

To analyse the development of discourse fields over time as the core component in register diversification, we trained a topic model with 30 topics⁷. Stop words were excluded and documents were split into parts of at most 5000 tokens each to control for largely varying document lengths.

Table 2 shows four of the 30 topics with their most typical words. Note that topics do not only capture the field of discourse (BIOLOGY 3) but also genre (REPORTING), mode (FORMULAE), or simply reoccurring boiler plate text (HEADMATTER).

Figure 5A displays the topic hierarchy resulting from clustering the topics based on the Pearson Distance between their topic-document distributions⁸. Labels for topics and topic clusters have been assigned manually, and redundant topics with very similar topic word distributions, such as BIOLOGY, have been numbered through.

Figure 5B shows the probabilities of the combined topics over time. As can be seen, the first hundred years are dominated by the rather generic combined topic REPORTING, which covers around 70% of the topic space. Indeed, the underlying topic REPORTING makes for more than 50% of the topic space during the first 50 years. Starting in 1750, topics become more diversified into individual disciplines, indicating register diversification in terms of discourse field. In addition, in line with the analysis in section 3.1, we clearly see the rise of the CHEMISTRY topic around the 1790s.

⁷For the corpus at hand, a smaller number of topics leads to conflated topics, a larger number to redundant topics.

⁸Clustering by Jensen-Shannon Divergence results in a less intuitive hierarchy.

TABLE 2 | Top five words for selected topics.

REPORTING	HEADMATTER	BIOLOGY 3	FORMULAE
great	vol	cells	equation
time	society	fig	equations
made	london	cell	function
found	author	tissue	form
account	part	nucleus	cos

As shown in **Figure 6A** diversification is evidenced by the clearly increasing entropy of the topic distribution over time. However, the mean entropy of the individual document-topic distributions remains remarkably stable, even though the mean number of authors per document and document length increase over time. Even the mean entropy weighted by document length (not shown) remains stable. This may be in part due to using asymmetric priors for the document-topic distributions, which generally skews them toward topics containing common words shared by many documents (Wallach et al., 2009), thus stabilizing the document-topic distributions over time.

Figure 6B shows the diachronic development of entropies at the level of words. The overall entropy of the unigram language model as well as the mean entropy of the topic word distributions weighted by the topic probabilities are also remarkably stable. However, the (unweighted) mean entropy of topic word distributions clearly increases over time. Indeed, due to the fairly high correlation of 0.81 (Spearman) between topic probability and the topic word entropy, evolving topics with increasing probability also increase in their word entropy, i.e., their vocabulary becomes more diverse. **Figure 7** demonstrates this for the evolving topics in the group LIFESCIENCE 2. All topics increase over time both

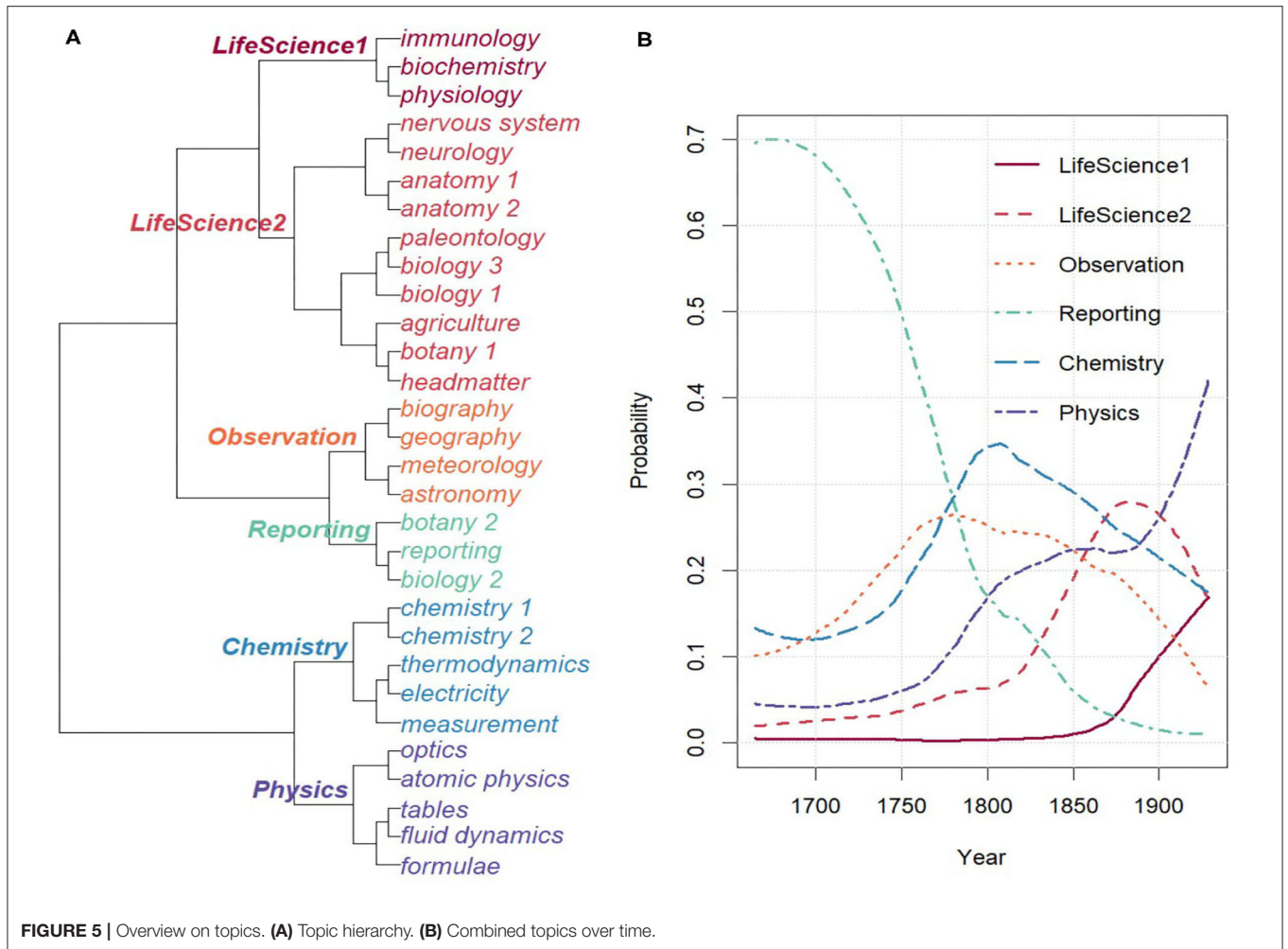


FIGURE 5 | Overview on topics. (A) Topic hierarchy. (B) Combined topics over time.

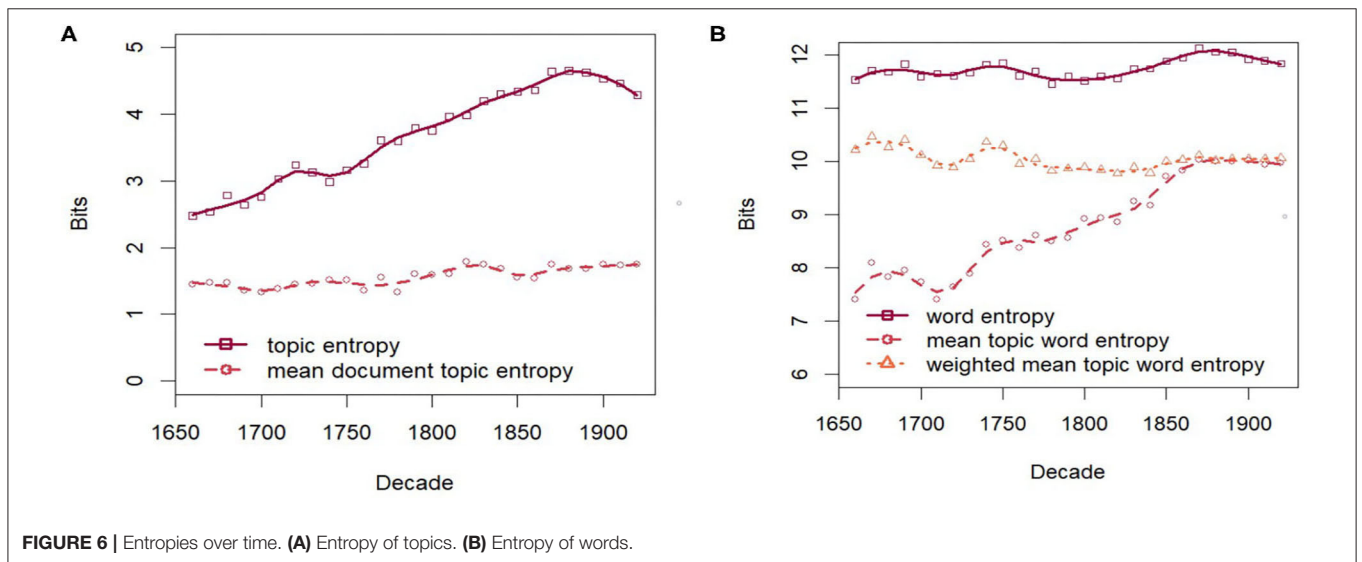
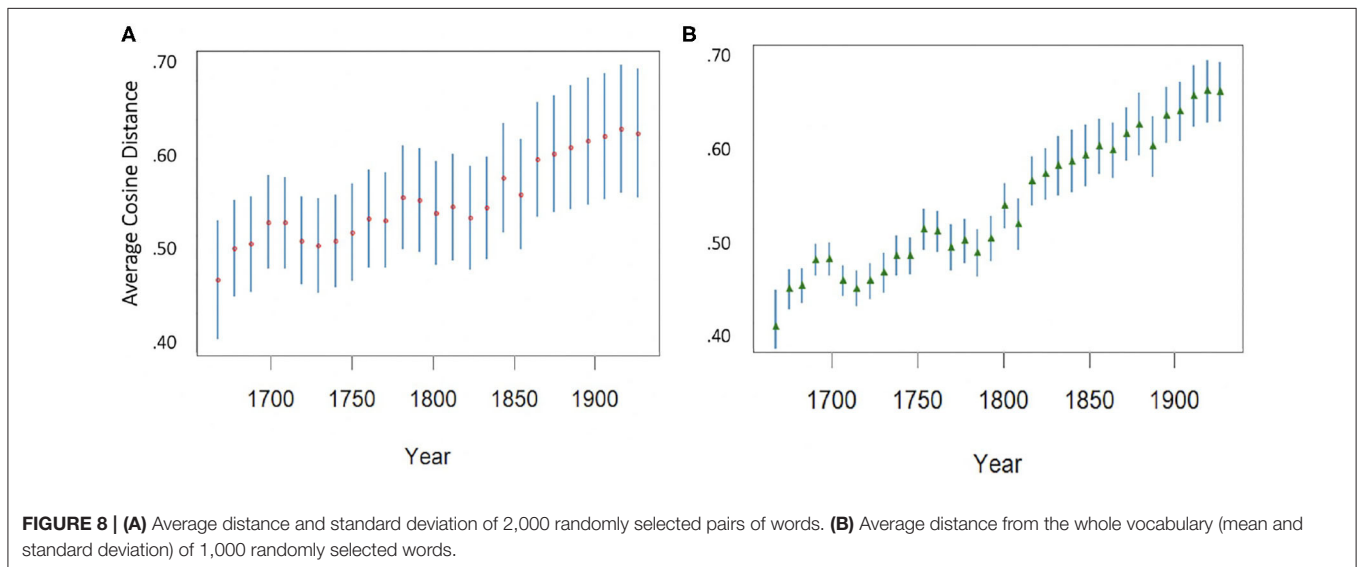
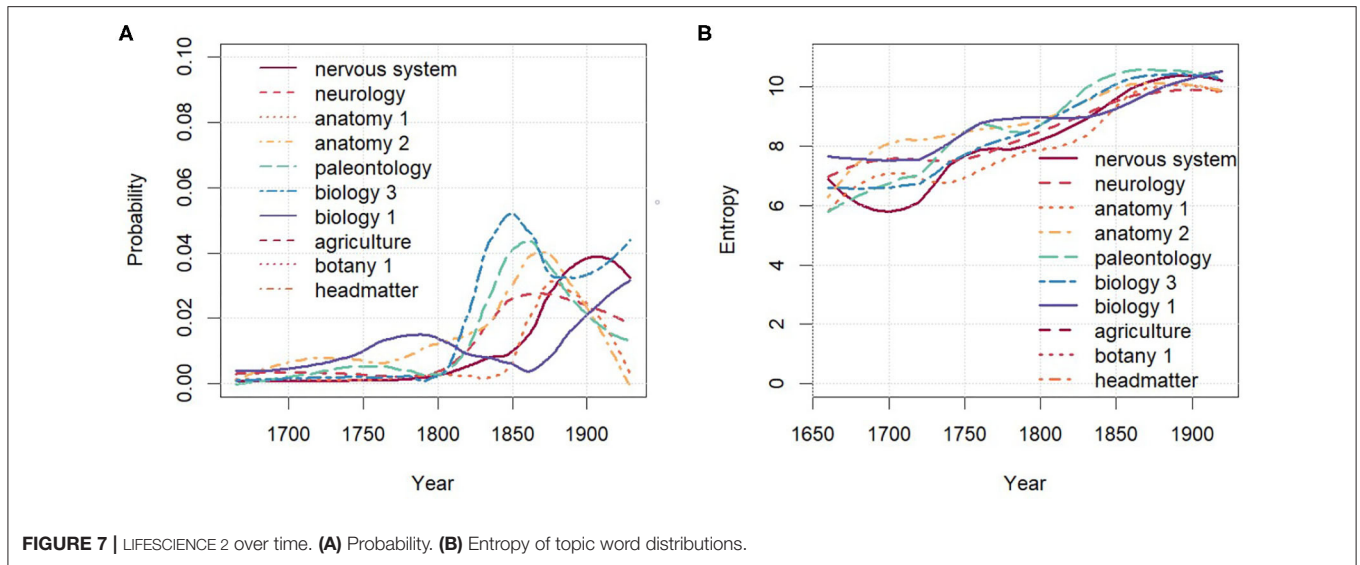


FIGURE 6 | Entropies over time. (A) Entropy of topics. (B) Entropy of words.



in probability and entropy⁹. As will be seen in section 4.3, this trend is mirrored in the analysis of paradigmatic word clusters by word embeddings.

4.3. Paradigmatic Effects

To gain insights into the paradigmatic effects of the diachronic trends detected by the preceding analyses, we need to consider word usage according to syntagmatic context. To capture grammatical aspects as well (rather than just lexical-semantic patterns), we take word forms rather than lemmas as a unit for modeling and we do not exclude function words.

Based on the word embedding model as shown in section 3.2.3, we observe that the word embedding space of

the RSC grows over time both in terms of *vocabulary size* and in terms of *average distance* between words. While a growing vocabulary can be interpreted in many ways, it is more informative to look at the increase in average distance between words. Here, not every term grows apart from all other terms (in fact, many pairs of words get closer through time) but when we take two random terms the average distance between them is likely to increase—see **Figure 8**: (A) shows the diachronic trend for the distance between 2,000 randomly selected pairs of words and (B) for the distance of 1,000 randomly selected words from the rest of the vocabulary. The words were selected among those terms that appear at least once in every decade. In both cases, the trend toward a growing distance is clearly visible.

Given that WEs are based on similarity in context, this means that overall, words are used increasingly in different contexts, a clear sign of diversification in language use. For example, the

⁹A similar correlation between probability and entropy can be observed in other rising topic groups.

usage of *magnify* and *glorify* diverges through the last centuries resulting in a meaning shift for *magnify* which becomes more associated with the aggrandizing effects of optical lenses while *glorify* remains closer to its original sense of elevating or making glorious. If we look for these two words in the WE space, what we see is, in fact, a progressive decrease of the distributional similarity between them: for example, in 1860 their cosine distance is 0.48, while in 1950 it has gone up to 0.62. The nature of their nearest neighbors also diverges: *magnify* increasingly shows specialized, optic-related neighborhoods (*blood-globule* in 1730, *object-lens* in 1780, *eyeglass* in 1810) while the neighbors of *glorify* remain more mixed (mainly specific but non-technical verbs, such as *bill*, *reread*, *ingratiate*, with low similarity). Finally, their movement with respect to originally close neighbors is also consistent: e.g., the distance between *glorify* and *exalt* does not change between 1860 and 1920, while *magnify* appears to move away and back toward *exalt* through the decades and is more than 25 degrees further from it in 1920 than in 1670 (from 0.45 to 0.70).

To provide another example, a similar evolution is apparent for *filling* and *saturating*: their distance grows from 0.37 in 1700 to 0.65 in 1920, a difference of almost 30 degrees. In the same lapse of time, the distance between *saturating* and *packing* goes from 0.27 to 0.70. Actually, the meaning of *saturating* was originally closer to that of *satisfying* and *packing*: its usage as a synonym of *imbuing*, and its technical sense in chemistry are more recent, and have progressively drawn the word's usage apart from that of *filling*.

As noted above, we observe an overall expansion of the WE space. To test whether this expansion is not a simple effect of the increase of frequency and number of words in each decade, we select a set of function words which exhibit stable frequency and should not change in usage over time (e.g., the functions of *the*, *and*, and *for* did not change in the period considered). If the expansion we observe is due to raw frequency effects, function words should drift apart from each other at a similar rate as content words. This appears not to be the case. As shown in **Table 3**, if we compare the group of function words to a group of randomly selected content words, such as verbs and nouns, we can see that the distances between the elements of such group grow much faster than the distances between function words. Purely functional words drift apart considerably less than words having a lexical meaning, indicating that the latter are probably causing most of the lexical expansion. Thus, words having a proper lexical meaning grow apart much faster *on average* than words having a purely functional role.

This behavior is not consistent with a raw frequency effect, or with the side effects of changes in the magnitude of training data. It looks like the distributional profile of words is, on average, growing more distinct in this specific corpus. And this does not happen only for new vocabulary, created *ad hoc* for specific contexts: even when we factor out the changes in lexicon and we consider only those words that appear in every decade (Persistent Vocabulary in **Table 3**), the effect is still visible. This interpretation is supported when we inspect the entropy on specific WE clusters over time. We consider two cases: increasing and decreasing entropy on a cluster, the former signaling lexical

diversification, the latter signaling converging linguistic usage. For instance, coming back to the field of chemistry, we observe increasing entropy in particular clusters of content words: see **Figure 9** for an example, showing (A) relative frequency of selected terms denoting chemical compounds and (B) entropy on the WE cluster containing those terms (radius of cosine similarity > 0.6).

As an example of the opposite trend, i.e., decreasing entropy, consider the use of *ing*-forms which diversify according to the analysis above shown for *filling* and *saturating*, i.e., they spread to different syntagmatic contexts. In the example in **Figure 10**, the terms in the cluster containing *assuming* exhibit a skewed frequency over time with decreasing entropy, reflecting in this case stylistic convergence, i.e., the tacit agreement on using particular linguistic forms rather than others. In particular, *assuming* has 30 close neighbors (including *supposing*, *assume*, *considering*) in the first decade, but only 13 close neighbors in the last decade, with *assuming*, *assume* dominating by frequency.

The effect of stylistic convergence on the reduction of the cluster entropy of *assuming* is visible also through a cursory look at some corpus concordances. Uses of *assuming* in the sense of “adopting” disappear (see example 1). Over time, *assuming* comes to be used increasingly at the beginning of sentences (example 2), the dominant use being the non-finite alternative to a conditional clause (*If we assume a/the/that...*). In terms of frequency, the dominant choice in the cluster is *assume*, presumably as a short form of *let us/let's assume* (example 3), a usage that is often associated with mathematical reasoning.

- (1) *No notice is taken of any effervescence or discharge of air while it was assuming this color* (Cavendish, 1786).
- (2) *Assuming a distribution of light of the form when x is the distance along the spectrum from the center of the line, the half breadth is defined as the distance in which the intensity is reduced to half the maximum* (Strutt, 1919).
- (3) *Assume any three points a, b, c in the surface, no two of which are on one generator, [...]* (Gardiner, 1867).

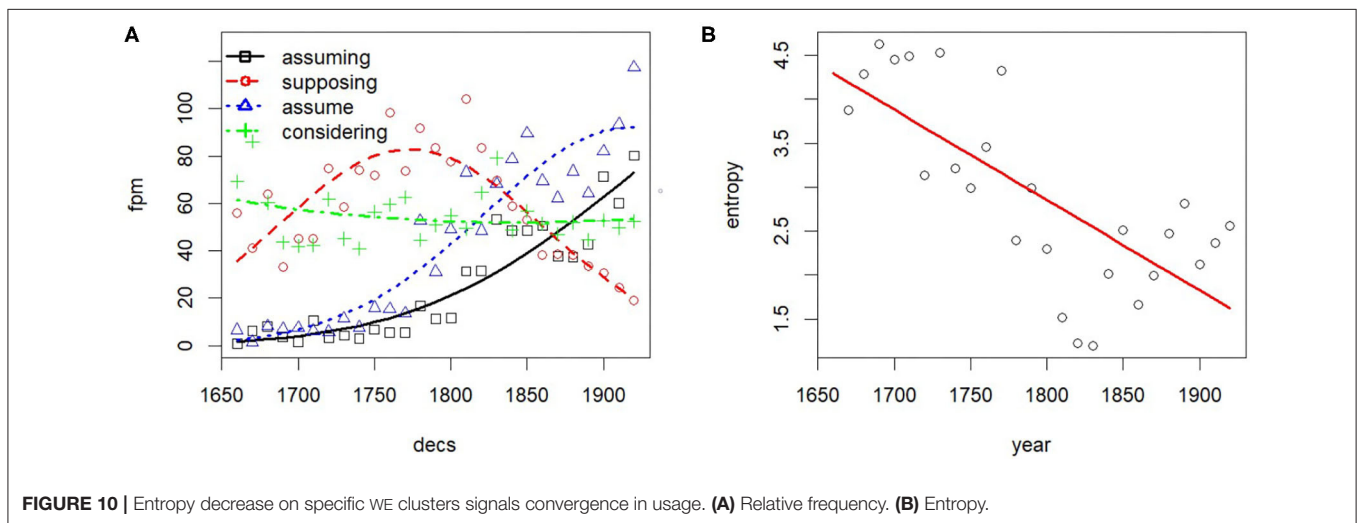
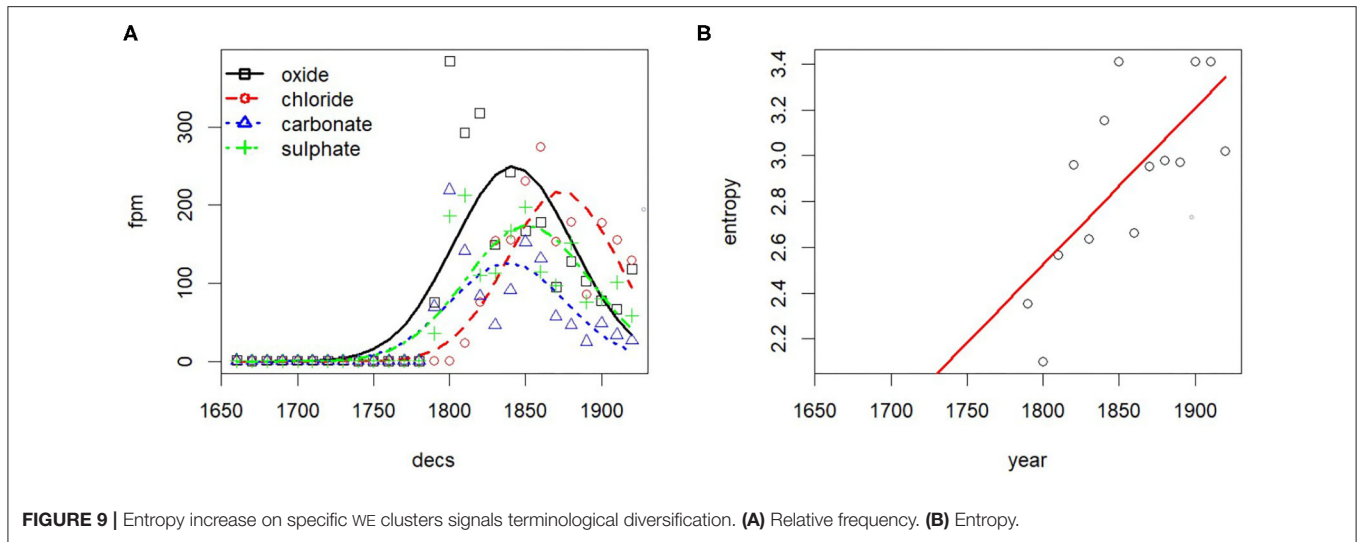
5. SUMMARY AND FUTURE WORK

We have explored patterns of variation and change in language use in Scientific English from a diachronic perspective, focusing on the Late Modern period. Our starting assumption was that we will find both traces of diversification in terms of

TABLE 3 | Average cosine distance between function words vs. 2,000 randomly selected content words in the first and last decade of RSC 6.0 Open.

Group	Full vocabulary		Persistent vocabulary	
	1670	1920	1670	1920
Function words	0.44	0.51	0.46	0.47
Content words	0.45	0.70	0.44	0.63

To account for the constantly updated vocabulary of scientific terminology, we present both the results for all words in each decade (Full Vocabulary) and for only those words that appear in every decade (Persistent Vocabulary).



discourse field, thus pointing to register formation, as well as convergence in linguistic usage as indicator of an emerging scientific style. As a data set we used 250+ years of publications of the Royal Society of London [Royal Society Corpus (RSC), Version 6.0 Open].

We have elaborated a data-driven approach using three kinds of computational language models that reveal different aspects of diachronic change. Ngram models (both lemma and POS-based) point to an overall trend of consolidation in linguistic usage. But the lexical level dynamically oscillates between high peaks marking lexical innovation and lows marking stable linguistic use, where the peaks typically reflect new scientific discoveries or insights. At the grammatical level, we observe similar tendencies but at a much lower level and rate and the consolidation trend is much more obvious. Inspecting the specific grammatical patterns involved, we find that they mark what we commonly refer to as “scientific style,” such as relational and passive clauses or specific nominal patterns for hosting terminology.

To investigate further the tendencies at the level of words, we have looked at aggregations of words from two perspectives—how words group together to form topics (development of fields of discourse as the core factor in register formation) and how specific words group together to form paradigms based on their use in similar contexts. Diversification is fully born out from both perspectives with glimpses of consolidation as well. Analysis on the basis of a diachronic topic model shows that topics diversify over time, indexed by increasing entropy over topic/word distributions, a clear signal of register formation. Analysis on the basis of diachronic word embeddings reveals that the overall paradigmatic organization of the vocabulary changes quite dramatically: the lexical space expands overall and it becomes more fragmented, the latter being a clear signal of diversification in word usage. Here, bursts of innovation are shown by increasing entropy on specific word clusters, such as terms for chemical compounds, mirroring the insights from lemma-based analysis with KLD. Also, patterns of convergence (confined uses of words) as well as obsolescence (word uses

leaving the language) are shown by decreasing entropy on particular word clusters, such as the cluster of *ing*-forms. Taken together, we encounter converging evidence of diversification at different levels of analysis; and at the same time we find signs of linguistic convergence as an overarching trend—an emerging tacit agreement on “how to say things”, a “scientific style.”

In terms of methods, we have elaborated a data-driven methodology for diachronic corpus comparison using state-of-the-art computational language models. To analyze and interpret model outputs, we have applied selected information-theoretic measures to diachronic comparison. Relative entropy used as a data-driven periodization technique provides insights into overall diachronic trends. Entropy provides a general measure of diversity and is applied here to capture diversification as well as converging language use for lexis (word embeddings) overall and discourse fields (topic models) in particular.

In future work, we will exploit more fully the results from topic modeling and the word embeddings model of the RSC. For instance, we want to systematically inspect high and low-entropy word embedding clusters to find more features marking expansion (vs. obsolescence) and diversification (vs. convergence). Also, annotating the corpus with topic labels from our diachronic topic model will allow us to investigate discipline-specific language use (e.g., chemistry) and contrast it with “general” scientific language (represented by the whole RSC) as well as study the life cycles of registers/sublanguages. Especially interesting from a sociocultural point of view would be to trace the spread of linguistic change across disciplines and authors (e.g., Did people adopt specific linguistic usages from famous scientists?). Finally, we would like to contextualize our findings from an evolutionary perspective and possibly devise predictive models of change. Our results seem to be in accordance not only with our intuitive understanding of the evolution of science but also with evolutionary studies on vocabulary formation

(e.g., Smith, 2004) showing how populations using specialized vocabularies are more likely to develop and take over when the selective ratio is pure efficacy in information exchange.

DATA AVAILABILITY STATEMENT

The Royal Society Corpus (RSC) 6.0 Open is available at: <https://hdl.handle.net/21.11119/0000-0004-8E37-F> (persistent handle). Word embedding models of the RSC with different parameter settings including visualization are available at: <http://corpora.ids-mannheim.de/openlab/diaviz1/description.html>.

AUTHOR CONTRIBUTIONS

YB curated the analyses on word embeddings showing the diachronic expansion of the scientific semantic space and lexical-semantic specialisation. SD-O carried out the analysis on diachronic trends in lexis and grammar using data-driven periodization and elaborating on features’ contribution to change. PF trained the word embeddings and the topic models and designed and implemented the entropy-based diachronic analysis. ET provided the historical background and was involved in hypothesis building and interpretation of results.

FUNDING

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 232722074 - SFB 1102. The authors also acknowledge support by the German Federal Ministry of Education and Research (BMBF) under grant CLARIN-D, the German Common Language Resources and Technology Infrastructure. Also, the authors are indebted to Dr. Louisiane Ferlier, digitisation project manager at Royal Society Publishing, for providing advice and access to sources.

REFERENCES

- Aitchison, J. (2017). “Psycholinguistic perspectives on language change,” in *The Handbook of Historical Linguistics*, eds D. Joseph and R. D. Janda (London, UK: Blackwell), 736–743. doi: 10.1002/9781405166201.ch25
- Aragamon, S. (2019). Register in computational language research. *Register Stud.* 1, 100–135. doi: 10.1075/rs.18015.arg
- Atkinson, D. (1999). *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675–1975*. New York, NY: Erlbaum. doi: 10.4324/9781410601704
- Banks, D. (2008). *The Development of Scientific Writing: Linguistic Features and Historical Context*. London; Oakville, OM: Equinox.
- Baron, A., and Rayson, P. (2008). “VARD 2: a tool for dealing with spelling variation in historical corpora,” in *Proceedings of the Postgraduate Conference in Corpus Linguistics* (Birmingham).
- Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511621024
- Biber, D., and Gray, B. (2011). “The historical shift of scientific academic prose in English towards less explicit styles of expression: writing without verbs,” in *Researching Specialized Languages*, eds V. Bathia, P. Sánchez, and P. Pérez-Paredes (Amsterdam: John Benjamins), 11–24. doi: 10.1075/scf.47.04bib
- Biber, D., and Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing. Studies in English Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511920776
- Bizzoni, Y., Degaetano-Ortlieb, S., Menzel, K., Krielke, P., and Teich, E. (2019a). “Grammar and meaning: analysing the topology of diachronic word embeddings,” in *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change* (Florence: Association for Computational Linguistics), 175–185. doi: 10.18653/v1/W19-4722
- Bizzoni, Y., Mosbach, M., Klakow, D., and Degaetano-Ortlieb, S. (2019b). “Some steps towards the generation of diachronic WordNets,” in *Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa’19)* (Turku: ACL).
- Blei, D. M., and Lafferty, J. D. (2006). “Dynamic topic models,” in *Proceedings of the 23rd International Conference on Machine Learning* (Pittsburgh, PA), 113–120. doi: 10.1145/1143844.1143859
- Bochkarev, V., Solovyev, V. D., and Wichmann, S. (2014). Universals versus historical contingencies in lexical evolution. *J. R. Soc. Interface* 11, 1–8. doi: 10.1098/rsif.2014.0841
- Burke, P. (2004). *Languages and Communities in Early Modern Europe*. Cambridge: CUP. doi: 10.1017/CBO9780511617362
- Bybee, J. (2007). *Frequency of Use and the Organization of Language*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195301571.001.0001
- Cavendish, H. (1786). XIII. An account of experiments made by Mr. John McNab, at Henley House, Hudson’s Bay, relating to freezing mixtures. *Phil. Trans. R. Soc.* 76, 241–272. doi: 10.1098/rstl.1786.0013

- Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., and Potts, C. (2013). "No country for old members: user lifecycle and linguistic change in online communities," in *Proceedings of the 22nd International World Wide Web Conference (WWW)* (Rio de Janeiro). doi: 10.1145/2488388.2488416
- Degaetano-Ortlieb, S., and Teich, E. (2016). "Information-based modeling of diachronic linguistic change: from typicality to productivity," in *Proceedings of the 10th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature at ACL2016*, 165–173. doi: 10.18653/v1/W16-2121
- Degaetano-Ortlieb, S., and Teich, E. (2018). "Using relative entropy for detection and analysis of periods of diachronic linguistic change," in *Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature at COLING2018* (Santa Fe, NM), 22–33.
- Degaetano-Ortlieb, S., and Teich, E. (2019). Toward an optimal code for communication: the case of scientific English. *Corpus Linguist. Linguist. Theory* 1–33. doi: 10.1515/clt-2018-0088. [Epub ahead of print].
- Dubossarsky, H., Hengchen, S., Tahmasebi, N., and Schlechtweg, D. (2019). "Time-out: temporal referencing for robust modeling of lexical semantic change," in *Proceedings of the 57th Meeting of the Association for Computational Linguistics (ACL2019)* (Florence: ACL), 457–470. doi: 10.18653/v1/P19-1044
- Dubossarsky, H., Weinshall, D., and Grossman, E. (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* (Copenhagen: Association for Computational Linguistics), 1136–1145. doi: 10.18653/v1/D17-1118
- Eisenstein, J., O'Connor, B., Smith, N. A., and Xing, E. P. (2014). Diffusion of lexical change in social media. *PLoS ONE* 9:e113114. doi: 10.1371/journal.pone.0113114
- Fankhauser, P., Knappen, J., and Teich, E. (2014). "Exploring and visualizing variation in language resources," in *Proceedings of the 9th Language Resources and Evaluation Conference (LREC)* (Reykjavik), 4125–4128.
- Fankhauser, P., Knappen, J., and Teich, E. (2016). "Topical diversification over time in the Royal Society Corpus," in *Proceedings of Digital Humanities (DH)* (Krakow).
- Fankhauser, P., and Kupietz, M. (2017). "Visual correlation for detecting patterns in language change," in *Visualisierungsprozesse in den Humanities. Linguistische Perspektiven auf Prägungen, Praktiken, Positionen (VisuHu 2017)* (Zürich).
- Fischer, S., Knappen, J., Menzel, K., and Teich, E. (2020). "The Royal Society Corpus 6.0. Providing 300+ years of scientific writing for humanistic study," in *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)* (Marseille).
- Fyfe, A., Squazzoni, F., Tornø, D., and Dondio, P. (2019). Managing the growth of peer review at the Royal Society journals, 1865–1965. *Sci. Technol. Human Values*. 45, 405–429. doi: 10.1177/0162243919862868
- Garcia, M., and García-Salido, M. (2019). "A method to automatically identify diachronic variation in collocations," in *Proceedings of the 1st Workshop on Computational Approaches to Historical Language Change* (Florence: ACL), 71–80. doi: 10.18653/v1/W19-4709
- Gardiner, M. (1867). *Memoir on Undevelopable Uniquadric Homographics*. [Abstract]. *Proc. R. Soc. Lond.* 16:389–398. Available online at: www.jstor.org/stable/112537
- Gleick, J. (2010). "At the beginning: More things in heaven and earth," in *Seeing Further. The Story of Science and The Royal Society*, ed B. Bryson (London, UK: Harper Press), 17–36.
- Goldberg, A. E. (2006). *Constructions at Work. The Nature of Generalizations in Language*. Oxford: OUP.
- Görlach, M. (2001). *Eighteenth-Century English*. Heidelberg: Winter.
- Görlach, M. (2004). *Text Types and the History of English*. Berlin, New York, NY: de Gruyter.
- Grieve, J., Nini, A., and Guo, D. (2016). Analyzing lexical emergence in Modern American English online. *Engl. Lang. Linguist.* 20, 1–29. doi: 10.1017/S1360674316000526
- Hall, D., Jurafsky, D., and Manning, C. D. (2008). "Studying the history of ideas using topic models," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (Honolulu, HI: Association for Computational Linguistics). doi: 10.3115/1613715.1613763
- Halliday, M. (1985a). *An Introduction to Functional Grammar*. London: Edward Arnold.
- Halliday, M. (1985b). *Written and Spoken Language*. Melbourne, VIC: Deakin University Press.
- Halliday, M. (1988). "On the language of physical science," in *Registers of Written English: Situational Factors and Linguistic Features*, ed M. Ghadessy (London: Pinter), 162–177.
- Halliday, M., and Martin, J. (1993). *Writing Science: Literacy and Discursive Power*. London: Falmer Press.
- Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2016). "Cultural shift or linguistic drift? Comparing two computational models of semantic change," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (Austin, TX). doi: 10.18653/v1/D16-1229
- Hardie, A. (2012). CQPweb – combining power, flexibility and usability in a corpus analysis tool. *Int. J. Corpus Linguist.* 17, 380–409. doi: 10.1075/ijcl.17.3.04har
- Harris, Z. S. (1954). Distributional structure. *Word* 10, 146–162. doi: 10.1080/00437956.1954.11659520
- Hilpert, M., and Perek, F. (2015). Meaning change in a petri dish: constructions, semantic vector spaces, and motion charts. *Linguist. Vanguard* 1, 339–350. doi: 10.1515/lingvan-2015-0013
- Hughes, J. M., Foti, N. J., Krakauer, D. C., and Rockmore, D. N. (2012). Quantitative patterns of stylistic influence in the evolution of literature. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7682–7686. doi: 10.1073/pnas.1115407109
- Hundt, M., Mollin, S., and Pfenninger, S. E. (eds). (2017). *The Changing English Language: Psycholinguistic Perspectives*. Cambridge, UK: CUP. doi: 10.1017/9781316091746
- Hunston, S., and Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: Benjamins. doi: 10.1075/scl.4
- Jucker, A. H., and Taavitsainen, I. (2013). *English Historical Pragmatics*. Edinburgh: Edinburgh University Press.
- Juola, P. (2003). The time course of language change. *Comput. Humanit.* 3, 77–96. doi: 10.1023/A:1021839220474
- Jurish, B. (2018). "Diachronic collocations, genre, and DiaCollo," in *Diachronic Corpora, Genre, and Language Change*, ed R. J. Whitt (Benjamins: Amsterdam), 41–64. doi: 10.1075/scl.85.03jur
- Kaiser, G. A., Butt, M., Kalouli, A.-L., Kehlbeck, R., Sevastjanova, R., and Kaiser, K. (2019). "Parhstvis: visualization of parallel multilingual historical data," in *Workshop on Computational Approaches to Historical Language Change* (Florence: Association for Computational Linguistics), 109–114.
- Kermes, H., Degaetano-Ortlieb, S., Khamis, A., Knappen, J., and Teich, E. (2016). "The Royal Society Corpus: from uncharted data to corpus," in *Proceedings of the 10th LREC* (Portoroz).
- Kim, Y., Chiu, Y.-I., Hanaki, K., Hegde, D., and Petrov, S. (2014). "Temporal analysis of language through neural language models," in *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science* (Baltimore, MD: Association for Computational Linguistics), 61–65. doi: 10.3115/v1/W14-2517
- Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141, 87–102. doi: 10.1016/j.cognition.2015.03.016
- Klingenstein, S., Hitchcock, T., and DeDeo, S. (2014). The civilizing process in London's Old Bailey. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9419–9424. doi: 10.1073/pnas.1405984111
- Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86. doi: 10.1214/aoms/117729694
- Kutuzov, A., Øvrelid, L., Szymanski, T., and Velldal, E. (2018). "Diachronic word embeddings and semantic shifts: a survey," in *Proceedings of the 27th International Conference on Computational Linguistics (Coling)* (Sante Fe, NM: ACL), 1384–1397.
- Leech, G., Hundt, M., Mair, C., and Smith, N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511642210
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Ital. J. Linguist.* 20, 1–31.

- Ling, W., Dyer, C., Black, A. W., and Trancoso, I. (2015). "Two/too simple adaptations of Word2Vec for syntax problems," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Denver, CO: Association for Computational Linguistics), 1299–1304. doi: 10.3115/v1/N15-1142
- McFarland, D. A., Ramage, D., Chuang, J., Heer, J., Manning, C. D., and Jurafsky, D. (2013). Differentiating language usage through topic models. *Poetics* 41, 607–625. doi: 10.1016/j.poetic.2013.06.004
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems* (Lake Tahoe, NV), 3111–3119.
- Moskovich, I., and Crespo, B. (eds.). (2012). *Astronomy Playne and Simple: The Writing of Science between 1700 and 1900*. Amsterdam: Philadelphia, PA: John Benjamins. doi: 10.1075/z.173
- Moxham, N., and Fyfe, A. (2018). The royal society and the prehistory of peer review, 1665-1965. *Historical J.* 61, 863–889. doi: 10.1017/S0018246X17000334
- Nevalainen, T. (2006). "Historical sociolinguistics and language change," in *Handbook of the History of English*, eds A. van Kemenade and B. Los (London, UK: Wiley-Blackwell), 558–588. doi: 10.1002/9780470757048.ch22
- Nevalainen, T., and Traugott, E. C. (eds.). (2012). *The Oxford Handbook of the History of English*. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780199922765.001.0001
- Nguyen, D., Dogruöz, A. S., Rosé, C. P., and de Jong, F. (2016). Computational sociolinguistics: a survey. *Comput. Linguist.* 42, 537–593. doi: 10.1162/COLI_a_00258
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Säily, T., Nurmi, A., Palander-Collin, M., and Auer, A. (eds.). (2017). "Exploring future paths for historical sociolinguistics," in *Advances in Historical Sociolinguistics* (Amsterdam: Benjamins). doi: 10.1075/ahs.7.01sai
- Schmid, H. (1995). "Improvements in Part-of-Speech Tagging with an application to German," in *Proceedings of the ACL SIGDAT-Workshop* (Kyoto).
- Schmidgen, H. (2011). *Das Labor/The Laboratory*. Europäische Geschichte Online/European History Online (EGO).
- Smith, K. (2004). The evolution of vocabulary. *J. Theoret. Biol.* 228, 127–142. doi: 10.1016/j.jtbi.2003.12.016
- Steyvers, M., and Griffiths, T. (2007). *Probabilistic Topic Models*. Hillsdale, NJ: Erlbaum.
- Strutt, R. J. (1919). Bakerian lecture: A study of the line spectrum of sodium as excited by fluorescence. *Proc. R. Soc. Lond. A* 96:272–286. doi: 10.1098/rspa.1919.0054
- Taavitsainen, I., and Hiltunen, T. (eds.). (2019). *Late Modern English Medical Texts: Writing Medicine in the Eighteenth Century*. Amsterdam: Benjamins. doi: 10.1075/z.221
- Taavitsainen, I., and Jucker, A. H. (eds.). (2003). *Diachronic Perspectives on Address Term Systems*. Amsterdam: Benjamins. doi: 10.1075/pbns.107
- Tahmasebi, N., Borin, L., and Jatowt, A. (2018). Survey of computational approaches to diachronic conceptual change. *arXiv[Preprint]a.rXiv:1811.06278*.
- Thoiron, P. (1986). Diversity index and entropy as measures of lexical richness. *Comput. Humanit.* 20, 197–202. doi: 10.1007/BF02404461
- Wallach, H. M., Mimno, D. M., and McCallum, A. (2009). "Rethinking LDA: why priors matter," in *Advances in Neural Information Processing Systems 22*, eds Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta (Vancouver, BC: Curran Associates, Inc.), 1973–1981.
- Xu, Y., and Kemp, C. (2015). "A computational evaluation of two laws of semantic change," in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci)* (Pasadena, CA).
- Yang, T.-I., Torget, A., and Mihalcea, R. (2011). "Topic modeling on historical newspapers," in *Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)* (Portland, OR).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bizzoni, Degaetano-Ortlieb, Fankhauser and Teich. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.