Check for updates

The evolutionary pattern of language in scientific writings: A case study of *Philosophical Transactions of Royal Society* (1665–1869)

Kun Sun[1,2] · Haitao Liu[2,3] · Wenxin Xiong[4]

Abstract

Scientific writings, as one essential part of human culture, have evolved over centuries into their current form. Knowing how scientific writings evolved is particularly helpful in understanding how trends in scientific culture developed. It also allows us to better understand how scientific culture was interwoven with human culture generally. The availability of massive digitized texts and the progress in computational technologies today provide us with a convenient and credible way to discern the evolutionary patterns in scientific writings by examining the diachronic linguistic changes. The linguistic changes in scientific writings reflect the genre shifts that took place with historical changes in science and scientific writings. This study investigates a general evolutionary linguistic pattern in scientific writings. It does so by merging two credible computational methods: relative entropy; word-embedding concreteness and imageability. It thus creates a novel quantitative methodology and applies this to the examination of diachronic changes in the *Philosophical Transactions of Royal Society* (PTRS, 1665–1869). The data from two computational approaches can be well mapped to support the argument that this journal followed the evolutionary trend of increasing professionalization and specialization. But it also shows that language use in this journal was greatly influenced by historical events and other sociocultural factors. This study, as a "culturomic" approach, demonstrates that the linguistic evolutionary patterns in scientific discourse have been interrupted by external factors even though this scientific discourse would likely have cumulatively developed into a professional and specialized genre. The approaches proposed by this study can make a great contribution to full-text analysis in scientometrics.

Keywords Scientific discourse · Professionalization · Information discrimination · Semantic similarity · Full-text analysis

✉ Kun Sun
   kun.sun@uni-tuebingen.de

✉ Haitao Liu
   lhtzju@yeah.net

Extended author information available on the last page of the article

🖄 Springer

Introduction

Discerning the patterns of cultural evolution is obviously crucial to gaining a better understanding of both human nature and how the world has formed and changed (Shennan 2009). The evolution of specific cultural phenomena can be investigated quantitatively with the development of computational technology and by drawing on the increasing amount of digitalized materials (Carr et al. 2017; Hutchison et al. 2018). These studies have helped to establish a quantitative science of cultural change and have also provided insights into the evolutionary patterns in human culture.

Science, which is an essential part of human culture and one of the most important factors in the making of modern society (Morris et al. 2013; Marks 1986), has undergone drastic changes in recent centuries (Jacob 1988; Kuhn 1970). Further, scientific publications are the most important means of recording and communicating developments in science (Lindsey 1978). In a manner parallel to the developmental route taken by science itself, scientific writings, after undergoing considerable changes in their structure and style over the course of centuries, have evolved into the current form (Biber and Gray 2016; Banks 2008; Atkinson 1998; Harmon 1989). A quantitative investigation of how scientific writings evolved is therefore enormously helpful in understanding the trends underlying the development of scientific culture.

Linguistic data contain an abundance of information on human activities and culture. It can be supposed that by quantitatively investigating the historical linguistic data, we can discern evolutionary patterns in human culture (Garg et al. 2018; Iliev et al. 2016; Michel et al. 2011). However, factors such as a lack of historical data or of significant historical events can hinder the discovery of the underlying patterns in human activities. Two crucial factors in finding significant evolutionary patterns in this domain are the availability of proper linguistic data and the mastery of effective computational methods. For instance, an important recent development is the use of linguistic changes in the available historical corpora as markers of social change (Hamilton et al. 2016b; Hills and Adelman 2015; Iliev et al. 2016).

However, studies such as those based on the calculation of frequency using diversified and heterogeneous texts in Google books (Michel et al. 2011) only show in a rough and noncontinuous fashion the impact of society on language use instead of discerning a consistent evolutionary pattern in a specific field. The evolutionary patterns in a specific human culture cannot be explored either segmentally through historical materials or by exhibiting simple correlations between linguistic changes and some historical changes. Instead, such patterns would ideally be specific, insightful and consistent, like those patterns found in biological evolution (Gould 2002; Shennan 2009). One of the most convenient and credible ways that are available currently for detecting the evolutionary patterns in scientific writings is the examination of diachronic linguistic changes. A fundamental supposition of this study is that it is essential to choose appropriate historical text materials (corpus) spanning historic periods in order to investigate specific and precise evolutionary patterns in a specific field. Consistent records of scientific development should provide the solid linguistic data for tracing the patterns of scientific development and the evolution of scientific culture.

Philosophical Transactions of the Royal Society (PTRS), which dates from 1665, is the world's first science journal. This journal has undergone considerable changes due to many different factors, such as changes in editors, in the topics of scientific studies, and in reviewing policies. However, notwithstanding these factors, the consistent publications

of this journal make it possible to trace by using the methods of linguistics, how science has developed in the past three hundred years. The formation of modern science occurred prior to the 20th century. The industrial revolution and the second industrial revolution also occurred in this period and were themselves influenced by developments in modern science. The period from the mid-17th century to the 19th century was clearly of great significance to human society. For these reasons, we selected the linguistic data from the PTRS for an investigation of evolutionary patterns in the diachronic development of scientific discourse.

It is acknowledged that, with the development in society and culture, the evolution of science exhibits a general tendency towards increasing professionalization and specialization (Casadevall and Fang 2014). This trend suggests that scientific writings, which record scientific achievements and communications, should also become increasingly professional and specialized (Houghton 1975; Ure 1982). Furthermore, according to a widely held belief, scientific discourse became more abstract, professional and specialized (Mack 2015; Gross et al. 2002; Martin and Veel 2005), differentiating itself from other genres. In past studies, the linguistic evolution of scientific discourse was mostly based on observation. However, some quantitative research on the linguistic evolution of scientific discourse was only based on the simple calculation of frequency of lexical and syntactic phenomena. Although *frequency* is very important and useful in quantitative research, it is just the sum of all the reasons why a word would occur often in a corpus. With the development of computational approaches, we have more methods such as that of word embeddings that actually implicitly measure aspects of meaning. The other problem with the aforementioned quantitative research is that it collected diversified and heterogeneous texts from many different scientific disciplines and sources (reports, journal articles, books, leaflets etc.).

The language in these texts was too wide-ranging as well as being influenced by numerous multifaceted, unknown and uncontrolled factors. All this increases the difficulty of quantitative research, diversifies the data and reduces its reliability. It is for this reason that consistent publications that are similar in form and that strive for the same long-term goal are easier to work, more amenable to quantitative research and the data therein is more reliable. Previous studies of historical records have focused on language use as indicative of larger shifts in style, grammar and usage. However, by focusing on a single journal that was published consistently, we can observe the interplay between the individual and the historical environment. More importantly, in the 17th, 18th and 19th centuries, scientific writings in journals might not have been as important as academic books or reports. However, journal papers are likely to play a crucial role in science communication at present. For this reason, it is interesting and worthwhile to see how the language style of writings in journals looked in earlier times and to see what changes such journal writings underwent. The PTRS furnishes perhaps the best material for investigating these issues. This is why it is the focus of our interest. Moreover, this field (diachronic study) has seldom been explored.

Scientific writings are an objective enterprise and their style is designed to a large degree to focus readers' minds on the scientific topics under consideration. Our specific concerns in the following are whether the evolution of language in the PTRS followed the general evolutionary trend of scientific discourse (towards professionalization and specialization) or whether it changed over time under the influence of events in the history of science and what external factors influenced the evolution of scientific writings. Our hypothesis is that this journal followed the evolutionary trend of becoming increasingly professionalized and specialized, as well as being influenced by external factors. Specifically, the hypothesis predicts that this evolutionary trend or route ought to be relatively linear. However, the straight evolutionary line could have

slight fluctuations caused by socio-cultural factors. Previous studies of early scientific journals and the PTRS demonstrate that the following factors might influence the development of a scientific journal. For instance, the academic society exerted a great influence on the quality and the goals of this journal (Atkinson 1998). More importantly, the professionalization of science as a career and of scientists forced scientific discourse to become professional and specialized in its language and structure through various strategies such as peer-reviewing (Dawson et al. 2020; Spier 2002). We propose a specific hypothesis concerning the influence of external factors that include the following elements: editors, sponsor, reviewing policy and science community. We hold that these external factors might have exerted a great influence on changes of language and style in scientific writings.

Additionally, the increasing availability of full text from scientific articles in machine readable electronic formats is an opportunity to greatly impact scientometrics. Glenisson et al. (2005) first combined full-text analysis and traditional bibliometric methods. Scientific writings by authors with different cultural background have different linguistic characteristics. Recent studies (e.g. Lu et al. 2019a, b; Chen et al. 2020) have shown that these different linguistic characteristics are also related to the characteristics of highly cited articles (i.e., impact) (e.g. Elgendi 2019). However, full-text analysis has seldom been used to make diachronic studies. These studies focusing on how linguistic complexity varies with the changes of scientific impacts share with the similar purpose of our study, that is, our study wants to know how language styles changed with authors who are clustered by different decades.

In order to understand these concerns and evaluate our hypothesis, we will adopt a new methodology to quantitatively investigate how language in this journal evolved over the period in question. Recently the following methods from computational linguistics have been extensively used in detecting language changes and other phenomena: relative entropy from information theory; linguistic concreteness and imageability based on word embeddings (Murdock et al. 2017; Garg et al. 2018; Snefjella et al. 2019). After relative entropy and word-embedding concreteness/imageability are merged into a new paradigm, the paradigm is able to detect linguistic changes comprehensively. This study will merge the two highly efficient computational methods (relative entropy and concreteness/imageability based on word embeddings) for cross-verification, thus creating a novel methodology for examining evolutionary patterns in the language of the PTRS. Our approach will potentially provide a new quantitative method in full-text analysis. The current study will address the following three questions:

(1) Did language in the PTRS evolve towards an increasing professionalization and specialization in the authors who are grouped together by decades?
(2) To what extent did language in the PTRS change over time under the influence of external socio-cultural factors?
(3) How much can relative entropy and linguistic concreteness contribute to quantitative methods in full-text analysis?

Background

As mentioned above, the evolutionary trend in scientific writings is towards professionalization and specialization (Houghton 1975; Ure 1982). Specialization means that the stylistic development of scientific writings moves towards a specialized genre because of the need for differentiation from other genres, as well as being partly due to the inauguration of different disciplines. Further, professionalization indicates that scientific writings, as means

of presenting and communicating science should become more efficient. The detection of linguistic changes is a reliable method for gaining insight into the trend of professionalization and specialization, as discussed in Introduction. In order to detect these changes, we need to choose the appropriate linguistic measures.

In earlier stages of linguistic investigation, *frequency* was an important measure for observing the evolution of language. For instance, Biber and Gray (2016) analyze historical changes in the grammatical complexity of academic writing by comparing their changes of frequencies. Their finding is that academic writings are often surprisingly imprecise, which was due to the loss of meaning that came with the use of compressed rather than elaborated formulations. Banks (2008) also adopts a similar frequency-based quantitative method to investigate the linguistic changes in the PTRS (1700–1980). One of his findings is that *nominalization* has been continuously increasing in the PTRS (1700–1980). Nominalization refers to the use of a word which is not a noun as a noun or as the head of a noun phrase, with or without morphological transformation. However, *frequency* does not consider the context. In order to reduce the limit of frequency, many studies attempted to use a diversity of methods to explore historically linguistic changes. For instance, *n-gram* can be used to calculate the bi-/tri-/four-gram transitional probability, which considers the larger context. Moreover, the *word vector* method can represent word meanings from a different perspective. That means that the relation between words can be measured by comparing their vectors. However, there are also other measures for investigating historical linguistic changes.

Measures in the information theory, such as relative entropy, can be used to evaluate diachronic changes in language and to detect cognitive evolution (Sherwin 2018). The use of relative entropy as a measure of changes in the probability distribution over linguistic features has proven effective in many studies (Murdock et al. 2017; Klingenstein et al. 2014; Barron et al. 2018). The other popular method for detecting linguistic changes often seeks association between neighboring words. For instance, comparisons between seed words and their word's immediate neighborhood (such as the lists of k nearest neighbors) can detect slight cultural shifts in word meaning (Kutuzov et al. 2018; Hamilton et al. 2016b). However, it is hard to ascertain which words are crucial in sciences that have a variety of greatly differing subfields. When sciences changed historically, it is much harder to confirm the crucial words. Further, word association or semantic similarity cannot be directly used to detect diachronic changes. A simple and efficient method for detecting semantic changes of words is linguistic concreteness. This method has already proved to be effective (Hills and Adelman 2015; Snefjella et al. 2019). On this approach, concreteness based on word embeddings provides a good measure for giving scores across different periods. Another measure, imageability, which is closely correlated with concreteness, is used in order to test the validity of linguistic concreteness. Word embeddings can be used for wide exploration; but they do not contain information on the probability of one word over different periods. Fortunately, relative entropy compensates for the weakness in word embeddings.

Relative entropy

In information theory, entropy (Shannon 1948) quantifies the amount of uncertainty involved in the value of a random variable or the outcome of a random process. "Relative entropy" or Kullback–Leibler divergence (KLD, also called "Bayesian surprise" or cognitive surprise) (Kullback and Leibler 1951), is derived from entropy. It refers to the number

of additional "bits" needed when a non-optimal encoding is used. Relative entropy is actually the expected discrimination information and it is used to measure the loss and gain of information. That is why it can measure diachronic changes in language. Formally, given two probability distributions $p(x)$ and $q(x)$ over a discrete random variable X, the relative entropy given by $D(p\|q)$ is defined as follows:

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} = \sum_{x \in X} p(x) * (\log p(x) - log q(x)) \tag{1}$$

$D(p\|q)$ (KLD) can also be understood as a measure of the information gained by revising one's beliefs from the prior probability distribution q to the posterior probability distribution p. KLD is therefore taken to measure cognitive surprise. When we talk about distribution information discrimination, KLD is closely associated with loss/gain of information. However, when we emphasize language users or readers cognitive experience, we speak of cognitive surprise (or Bayesian surprise). Here p typically represents the "true" distribution of data or observations, while p typically represents an ideal data or theoretic data. The order of p and q in (1) cannot be reversed. Note that KLD is applied over the same underlying set of events.

A recent strand of data-driven approaches in the analysis of diachronic change applies KLD measures. For instance, Murdock et al. (2017) applied KLD to find that there is a correlation between the major intellectual periods in Darwin's career as identified both by scholarship and his own self-commentary in his reading notes. Barron et al. (2018), using KLD between sequential speeches, found the contours of a new rhetorical space in the French Revolution. KLD was used to analyze the linguistic development of scientific writing (the PTRS) over time in order to discern change specific to scientific writings (Degaetano-Ortlieb and Teich 2018).

We can use the KLD method of a sliding window (e.g. 10 years) to detect changes in units of language. Specifically, KLD is used to compare preceding (pre period) and subsequent (post period) years in the sliding window. We can obtain Eq. (2) based on Eq. (1):

$$\begin{aligned} D(\text{post}||\text{pre}) &= \sum_i p(\text{unit}_i|\text{post})(\log_2 p(\text{unit}_i|\text{post}) - \log_2 p(\text{unit}_i|\text{pre})) \\ &= \sum_i p(\text{unit}_i|\text{post}) * \log_2 p(\text{unit}_i|\text{post}) \\ &\quad - \sum_i p(\text{unit}_i|\text{post}) * \log_2 p(\text{unit}_i|\text{pre}) \end{aligned} \tag{2}$$

Diachronic word concreteness

Words can roughly be classified into two groups: concrete and abstract. Concrete words usually denote relatively concrete entities, events, or actions. In contrast, abstract words denote feelings, ideas, social concepts, and introspective states. Generally, *concreteness* evaluates the degree to which the concept denoted by a word refers to a perceptible entity. Numerous studies have tried to examine the effect of concreteness and abstractness on psycholinguistics and cognitive studies, including cognitive processing, effects in working memory, storage and retrieval (Brysbaert et al. 2014; Hill et al. 2014). Concreteness and abstractness have been successfully used to detect linguistic diachronic changes (Hills and Adelman 2015; Snefjella et al. 2019). Changes of concreteness definitely reflect a tendency

towards using abstract, difficult diction and terms in scientific writings. Such changes, or the lack therefore, can also show whether articles in the PTRS became more professional and specialized or not. Furthermore, although concreteness is the main measure, another dimension, that of *imageability*, will be measured so as to allow comparisons and so to test the validity of concreteness. *Imageability* represents the degree of effort involved in generating a mental image of something (imageable, unimageable). The core textual items are likely to be made up of highly imageable words. Although imageability is strongly correlated with concreteness (Paivio et al. 1968), the two have seldom been used to capture distinct semantic aspects of a word (Scott et al. 2019). In a similar vein, changes of imageability also reflect a tendency towards abstraction in scientific discourse. When the degree of imageability is lower, this means that more difficult and less imageable words are used. The changes of both concreteness and imageability can show whether the PTRS became increasingly professional and specialized or not.

Quantitative use relies on the concreteness and imageability norms. The recent collection of large-scale concreteness norms has been provided by online participants who ranked concreteness (Brysbaert et al. 2014). For instance, concreteness ratings in the database of Brysbaert et al. (2014) were obtained from over 4000 participants by means of collecting subjective norms and it also used the Amazon Mechanical Turk for data collection, all of whom self-identified as native speakers of American English residing in the U.S. Each participant was given one or more lists of words and was asked to rate each word using a 5-point rating scale going from most abstract to most concrete. The database collection of imageability (Scott et al. 2019) was done online via an in-house experimental platform by 100 native English speakers from the University of Glasgow.

However, the method for assessing the concreteness of an individual word is fairly distinct from the method for measuring the degree of concreteness of a corpus (text/document). To date, there have been the three main methods used to assess concreteness degree of a corpus (text/document). The first method involves searching for the words (lemmatized tokens) in a corpus and then retrieving the rating norms for these words in the concreteness database and finally calculating the mean of all rating norms of these words in this corpus. The second method proposed by Hills and Adelman (2015) estimates the concreteness of a word according to the word's concreteness rating as weighted by the word's frequency of occurrence in a corpus. The third method by Hamilton et al. (2016a) and Snefjella et al. (2019) is to measure the semantic similarity between "seed words" and all words in this corpus in order to calculate the concreteness degree of each word in a corpus. The following compares these three methods.

Using the first method, the Coh-Metrix tool (Graesser et al. 2004) was created to calculate the degree of word concreteness (and the other word properties) for the content words in a corpus. Another tool for Automatic Analysis of Lexical Sophistication (TAALES) (Kyle et al. 2018) was used to compute lexical features related to word concreteness and it employed a similar method to Coh-Metrix. However, this approach is potentially problematic with regard to this mean because the degree of concreteness for a corpus is not equal to the mean of the summation of the scores of all (unique) content words in this corpus. The first possible deficiency concerns the frequency with which a word is considered in this method. The description of word frequency cannot be found in the relevant literature on the two tools, which strongly suggests that word frequency was not considered in this method. Secondly, even if all the content words were included, the final score is still potentially problematic. This be because the status of the remaining words in this corpus is unknown. The degree of given concreteness in a corpus must consider all the words in this corpus, including function words. For example, Brysbaert et al. (2014) provided the database of

concreteness norms for function words such as numbers, pronouns, determiners, propositions and the other types of function words. Overall, this method is hardly capable of precisely measuring the concreteness of a corpus.

The second method by Hills and Adelman (2015) assessed the linguistic concreteness of historical changes in a corpus according to a word's concreteness ratings as weighted by the word's frequency of occurrence in this corpus. The method is potentially better than the first method for summing up the concreteness scores of words in a corpus. The reason for this is that the method in Hills and Adelman (2015) solved the problem of neglecting words that occur twice or more. However, if there had been an abrupt change in the meanings of many words or a periodic fluctuation in meanings of many words over time, this method would not have been able to account for this. In this sense, the second method is a "static" approach to evaluating historical changes.

Before reviewing the third method, we need to introduce one crucial concept. The meanings of words can be represented using vectors, as part of a high-dimensional "semantic space", and these vectors are called "word embeddings". i.e. word embeddings are numbers which are converted from texts (or corpus). Word embeddings can represent co-occurrences of words in a corpus, and each vector represents the typical context for one word. The currently popular algorithm is the word2vec (Mikolov et al. 2013). This is a program that is especially predictive with regard to learning word embeddings from raw text. In this sense, word embeddings contain the semantic information on word meanings, word frequency, word meaning and context etc. In addition to this, word embeddings have been widely applied in natural language processing, cognitive sciences and language studies (ref. Bakarov 2018). Word embeddings are also widely used to explore diachronic changes in language and culture (Hamilton et al. 2016b; Garg et al. 2018). However, these studies are based on word association or semantic similarity. A key problem in the current context that must be taken into consideration is that of ascertaining which words are crucial in sciences that have a variety of subfields. Clearly, it is very hard to ascertain the crucial words when we, as in this study, are exploring in a variety of disciplines spanning two centuries. Instead of using word association, an examination of the change of word concreteness is a simple and reliable way of detecting linguistic changes.

The advantage of word embeddings has fostered the third method for calculating word concreteness. Hamilton et al. (2016a) proposed a new method for quantifying language changes based on the notion of word embeddings. Snefjella et al. (2019) borrowed the algorithms of Hamilton et al. (2016a, b) to measure the diachronic changes of linguistic concreteness. If the method of Hills and Adelman (2015) is "static", the method of Hamilton et al. (2016b) is "dynamic". The "dynamic" method of Snefjella et al. (2019) is potentially better than the "static" one. The current study explores the possibility of extending this dynamic method to the assessment of diachronic concreteness and imageability among historical sub-corpora of the PTRS.

The relation between the two methods

As mentioned above, either relative entropy or concreteness/imageability has frequently been used to detect language changes and to discern the evolutionary pattern in language changes. The current study will merge these two types of methods because of their unique strengths in detecting language changes. This will allow us to discern whether or not language in the PTRS evolved towards an increasing professionalization and specialization.

The second reason for the merger is that the two methods can cross-verify with each other, thus increasing the validity and reliability of the results.

In addition to the advantages of the convergence of the different measures, there is another substantial benefit that arises from the merging of the two methods, i.e., the two methods can be merged because these methods complement each other. According to Saussure's system, word function is dependent on a large degree on knowledge of both the syntagmatic and paradigmatic structure of language (De Saussure 2011). A syntagmatic relation is a type of semantic relation between words that co-occur in the same sentence or text. A paradigmatic relation is a different type of semantic relation between such words that can be substituted with another word in the same category. Syntagmatic relations allow probabilistic predictions. KLD gives the actually expected discrimination information and this is used to measure the loss and gain of information for the same set of linguistic units. The information discrimination (KLD) for a set of linguistic units over two periods is really helpful in discerning whether language has evolved towards a high degree of difficulty for language users or not. In this sense, the KLD approach can calculate the difference of the information distribution concerning one linguistic unit over two periods, which can measure language knowledge from the perspectives of both language communication and syntagmatic relation. By contrast, the paradigmatic structure captures semantic similarity. Word embeddings can be applied to measure word semantics which latter is essentially a kind of paradigmatic relation. The measure of concreteness/imageability based on word embeddings take advantage of semantic similarity (paradigmatic relation) is capable of incorporating the word semantics on the paradigmatic level from the perspectives of semantics and paradigmatic relation.

Moreover, as mentioned above, relative entropy is capable of measuring whether or not language has become more difficult for language users to process from the perspective of cognitive surprise. And as we know, diachronic concreteness/imageability is also able to estimate whether language has evolved towards a higher degree of difficulty/abstraction or not. This means that the two methods can be merged to detect whether the language has become more difficult and abstract or not. This is also the fourth reason that the two computational methods were merged to form a new paradigm.

When the two methods are applied to conceptually distinct relations that involve different dimensions of linguistic structure, an integrated approach allows the exploration of diachronic linguistic changes in PTRS in a comprehensive manner. The diachronic trend in scientific discourse can be represented in linguistic data and therefore can be detected through measuring the changes of relative entropy and linguistic concreteness/imageability. After relative entropy and word-embedding concreteness/imageability are merged into a new paradigm, the paradigm is able to detect linguistic changes comprehensively. In this sense, this study will create a novel methodology for examining evolutionary patterns of language.

Full-text analysis in scientometrics

The increasing availability of full text from scientific articles provides more methods in scientometrics. For instance, Glenisson et al. (2005) first combined full-text analysis and traditional bibliometric methods. In-text citations and entity metrics are typical examples of full-text analysis in scientometrics (Ding et al. 2013). However, it will be effective to analyze the other metrics analyzing full-texts. Full text contains additional information that

has not been available in bibliographic data. Full text also contains a relatively high level of detail about motivation, methods, data, instruments, results, and conclusions that authors typically report when documenting and submitting their work for publication (Boyack et al. 2013).

Full-text analysis in scientometrics mainly focused on in-text citations, such as reference position (e.g. Boyack et al. 2018), citation contexts or sentiments (e.g. Liu and Chen 2013; Ding et al. 2014; Lu et al. 2017), entity metrics (e.g. Ding et al. 2013; Mckeown et al. 2016), etc. Recently, linguistic complexity of scientific writing styles and scientific impacts (e.g. Lu et al. 2019a, b; Chen et al. 2020) have been studied to connect with characteristics of a highly cited article (e.g. Elgendi 2019) and author cultural background (Lu et al. 2019a, b).

However, full-text analysis has seldom been used to explore in changes in linguistic phenomena over time (i.e., diachronic analysis), that is, most full-text studies focused on synchronic studies. These studies on full-text analysis focused on how linguistic complexity varies with the changes of scientific impacts or authors cultural background. As a matter of fact, the present study has some similarity with these studies of linguistic complexity to some degree. As a matter of fact, our study uses similar criteria for classifying authors. The present study seeks to find out how language styles changed with authors who are grouped together in terms of decades. How language styles of scientific writings in journals changed in authors who are clustered by different decades. This will be quite useful in better understanding how the scientific literature evolved. Further, the scientific writings in the PTRS (1665–1869) did not use in-text citations and did not include the modern style abstracts and keywords. In this case, there have been no databases on bibliometric factors regarding the PTRS. Without the databases containing bibliometric factors on the PTRS, the traditional scientometrics methods seem to be difficult to implement in the present study. Despite this, seeking to understand how writing style evolved in scientific journals is consistent with a broadly-defined scientometrics approach which concerns itself with measuring and analyzing scientific literature.

Materials and methods

Materials

The PTRS (1665–1869) can be taken as representative of scientific journals in 17th, 18th and even early 19th centuries. Moreover, there were few scientific journals before 1800s. Journals play a uniquely important role in the history of science communication, so for this reason alone they are worthy of research. It is interesting and worthwhile to see how the writing in journals looked in earlier times and to see what changes such journal writings underwent. Consistent records of scientific development should provide solid linguistic data for tracing the patterns of scientific development and the evolution of scientific culture. The PTRS furnishes perhaps the best material for investigating these issues.

This study made use of The Royal Society Corpus (RSC) which collected all papers published in the PTRS from 1665 to 1869. The RSC materials are in a well-formed XML format (cf. vrt-format) including meta-data (e.g. author(s), text type, year of publication, text ID, and title, etc.) and ultimately form a corpus with 35 million tokens (cf. Kermes et al. 2016). The corpus is tokenized and linguistically annotated for lemma and part-of-speech (POS) using TreeTagger (Schmid 1999). In the vrt-format (vertical text format)

annotations on the token level (positional attributes, e.g. word, pos, lemma) are represented one-word-per-line with TAB delineated columns for each positional attribute. Annotations beyond the token level (structural attributes, e.g. texts, sentences, pages) are represented as SGML-tags with possible attribute-value pairs. Metadata are encoded as attributes of the *<text>-element*. (More information on the annotation of the corpus is available at https://fedora.clarin-d.uni-saarland.de/rsc/annotation.html.)

The following discusses why the PTRS published before 1880s was chosen. In 1887, the PTRS split into series "A" and "B", dealing with the physical and biological sciences respectively, afterwards into still further series again. From 1893, the *Proceedings,* another journal which is similar to the PTRS, began to include full research papers in its own right. The *Proceedings* split in two in 1905. *Notes and records* was established in 1938. The Royal Society later launched even more journals. The PTRS was divided into several sections and each section began to become more specialized. However, these series replaced the treatment of specific issues in the early PTRS. After the 1880s, the PTRS became very complicated. In short, it was no longer a single comprehensive journal. The PTRS published before 1880s is different from its succeeding series. This indicates that after 1880s the PTRS should not be treated as being the same journal that was published under this name prior to the 1880s.

The RSC contains all documents from this journal (1665–1869)l. These consist of four types of articles. But because book reviews, obituaries and abstracts are not real research papers, the present study only uses the full-length articles as its material. The total number of full-length papers is 7283, and the number of tokens per paper is 3223. After downloading RSC, we processed the corpus carefully and only extracted the texts of full-length articles. In terms of the tag of "decade of publication" in the RSC, we divided the whole corpus into twenty-one sections according to each decade. Twenty-one sub-corpora were therefore obtained. The following example illustrates the division. The article published in the year of 1810 was annotated as belonging to the decade "1810". Another article published in the year 1815 was also annotated as belonging to the decade "1810". That means that full-length articles published between 1810 and 1819 are included in the decade "1810". The two articles are therefore classified into the sub-corpus of "1810s". This classification also suggests that the corpus is split with authors who are grouped by different decades. The following table is the basic information on the PTRS that was processed in this study. In Table 1, "tokens" means all words, while "types" refers to the lemma. For example, "ran", "run", "running" are treated as one lemma (type), but they are three different tokens. "Embeddings" refers to words that have vectors processed by FastText which will be discussed below. The four items in Table 1 shows an increase in historical frequency.

Methods

Although both methods were implemented in these twenty-two sub-corpora, each method is used independently. We need to extract some "linguistic units" from each sub-corpora with respect to "relative entropy". In contrast, concreteness/imageability requires transforming each sub-corpus (texts) into word embeddings so as to compute the semantic similarity between each word and the "seed word". The road map for this study is shown in Fig. 1.

Table 1 The size for each sub-corpus of PTRS

Decade	Article number	Token in RSC (m)	Token (m)	Type	Embedding
1660s	270	0.46	0.30	16.9k	4232
1670s	291	0.83	0.45	20.91ek	5574
1680s	225	0.58	0.38	20.5k	5197
1690s	328	0.73	0.49	25.7k	6315
1700s	331	0.79	0.57	26.9k	6621
1710s	180	0.50	0.36	22.2k	5461
1720s	284	0.54	0.42	19.6k	5384
1730s	342	0.62	0.47	20.3k	5951
1740s	489	1.01	0.80	28.1k	8847
1750s	483	1.21	0.95	31.4k	9421
1760s	407	1.00	0.75	25k	7769
1770s	408	1.59	1.2	31.4k	9955
1780s	300	1.50	1.12	23k	8256
1790s	219	1.50	1.11	47.3k	8805
1800s	221	1.70	1.11	31.4k	8287
1810s	260	1.50	0.98	27.3k	7865
1820s	286	2.03	1.26	30k	8850
1830s	283	3.02	1.9	42.1k	11,415
1840s	216	2.50	1.52	38.7k	10,506
1850s	630	5.13	3.29	63.4k	15,717
1860s	730	6.59	4.04	77k	18,185
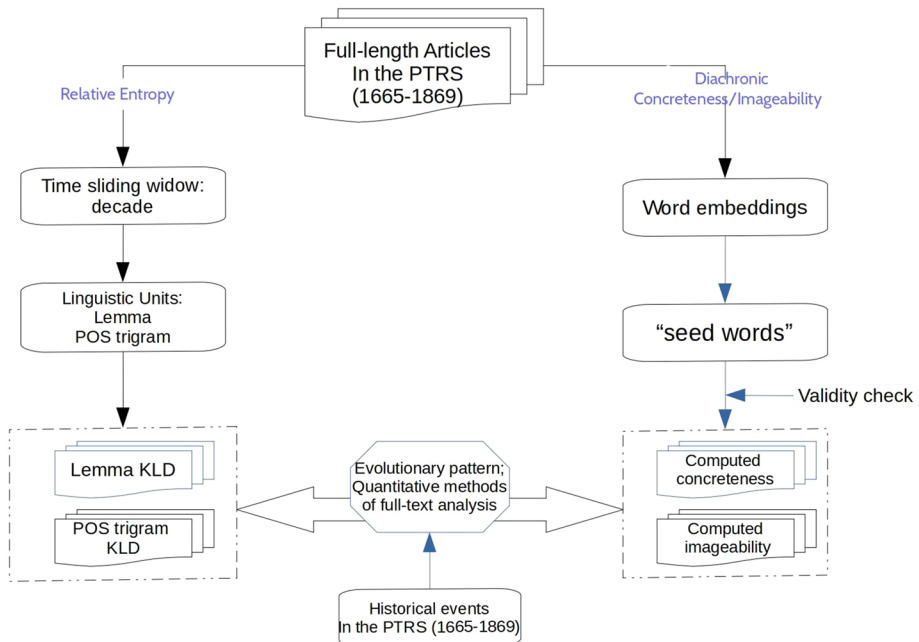
Fig. 1 Road map for this study

Relative entropy: lemma and POS trigram

As we know, KLD is used to compare preceding (pre period) and subsequent (post period) years in the sliding window. As shown in Eq. (2), there are three components: *unit, post and pre*. "Post" and "pre" represent the preceding and subsequent periods respectively, whereas "unit" here refers to linguistic units. In order to compute relative entropy, we have to specify the periods and linguistic units.

Here "pre" might have two cases: only some of the previous years (e.g. a decade); the average of all previous years. "Post" can be understood in the same way. We use the KLD method of 10 years to detect changes in units of language. A decade unit was chosen because scientific progress and the corresponding linguistic changes can be clearly seen over the period of a decade. Further the journal had existed for two hundred years prior to 1869, which divides into twenty-two decade units. For instance, if "1800s" is a "pre" period, "1810s" is a "post" period.

Relative entropy (KLD) in Eq. (2) measures the average amount of additional bits per linguistic unit needed to encode the same linguistic unit distributed according to "post" by using an encoding optimized for "pre". Applied to the comparison of sub-corpora of PTRS, KLD serves as a strong indication of the degree of difference between two sub-corpora (representing two different periods) measured in bits as well as the linguistic units that are primarily associated with a difference, That is to say, the difference of KLD indicates that linguistic units need high amounts of additional bits for encoding. After sliding over the time line of the PTRS comparing adjacent time periods (one decade) we can find KLD as indicators of change.

"Linguistic units"

The other concern here is that of the options of the "linguistic unit". When we speak of a linguistic unit in this study, the term here is confined to "lemma" and "POS trigram" for their typicality. The "lemma" (i.e. uni-gram form) refers to the canonical form, or dictionary form of a set of words. For instance, "take" is the lemma of "takes", "taking", "took" and "taken". On the other hand, the POS trigram here refers to a bundle consisting of three words marked by parts of speech. e.g. "the quantity of" is a trigram and the trigram's POS (part of speech) is Determiner-NOUN-Preposition (abbreviated as "DT-NN-IN").[1] According to Jurafsky and Martin (2008: 24), it is more common in practice to use trigram models in the field of natural language processing, because a trigram model depends on the previous two words rather than the previous word. Additionally, according to Biber et al. (1999: 994–995), 3-word lexical bundles have a much higher frequency than 4-word or 5-word lexical bundle. Although the scope of lexical bundles is smaller than *n*-gram, its frequency can still reflect users' preferences. Because of these, we use the trigram POS to present grammar and detect its changes. We wish to obtain those specific lemmas or POS trigrams that make the largest contribution in pre period. This allows us to differentiate the post period.

In order to work efficiently with relative entropy, a frequency difference between "pre" (a decade) and "post" (a subsequent decade) periods and *p* value in *Welch's t* test are used

---

[1] Part-of-speech tags are accordance with Penn Treebank Tagset. This is available at https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html.

to select the lemmas or the POS trigrams that are involved in change. It is noted that all frequencies are standard ones, based on one million words. The frequency of POS trigrams is not as great as that of lemmas. Accordingly, we employ two different procedures so as to make them work efficiently.

With respect to lemmas, all strings containing digits or symbols were removed in order to preserve pure texts for each decade. After 119 common stopwords[2] were removed from lemmas in each sub-corpus, the lemmas were filtered again by choosing these lemmas with a length greater than 2. Afterwards, lemma candidates were filtered again through the *Welch's t* test to determine significant differences between the relative frequencies of these units in a preceding decade and its subsequent decade (p value < 0.01). After the selection, each pair of consecutive decades has the same vocabulary for calculating the KLD.

With regard to the POS trigram, after obtaining all forms of POS trigram, we deleted those strings containing digits, punctuation and special marks (e.g. the mark of sentence ending, SENT, SYM). The most frequent and common POS trigram patterns were identified over the course of twenty-two decades and 87 POS trigram patterns were treated as "stopwords".[3] After these stopword-style POS trigrams were removed, we made use of the two measures for selecting those POS trigrams that help to calculate the KLD between one decade: the frequency difference and *Welch's t test* result (p value < 0.01). Similar to the lemma, after the selection, each pair of consecutive decades has the same repository of the POS trigram for calculating KLD.

As mentioned in "Background" section, KLD can be used to measure cognitive surprise (or Bayesian surprise). Specifically, KLD can quantify cognitive surprise through detecting distinctions in information between two periods (Murdock et al. 2017; Sayood 2018; Itti and Baldi 2009) and a high KLD indicates linguistic novelty in comparison with the past. KLD is also associated with a change in the reader's reactions when encountering the unexpected. It is therefore used to examine the evolution of cognitive surprise in the PTRS. In addition to giving an account of the discrimination of information distribution, "pre" and "post" in the equation above can be understood as cognitive surprise. That means the "pre" is the distribution of linguistic phenomena that readers have encountered before and "post" is the new distribution

---

[2] In natural language processing, stop words are the most common words in a language. These words are usually regarded as less useful in most natural language processing tasks. These stopwords are mostly short function words, such as the, is, at, which, and so on. There is no single universal list of stop words available. The length of different stopword lists vary greatly. Here we choose a short list that is widely used in natural language processing and that is available at https://www.textfixer.com/tutorials/common-english-words.txt.

[3] The most frequent POS trigrams among the twenty-two sub-corpora have been chosen. "DT-NN-IN" is one of them. Just like "stopwords", these trigram POS patterns are commonly used over this period and they are not significantly discriminative from the perspective of trigrams.

that readers will encounter.[4] When KLD tends to become larger over time, it suggests that language users find it more difficult to process these language units over this period.

Dynamic approach to concreteness/imageability

We have reviewed the strengths of the "dynamic" approach. We will use this method to compute the degree of concreteness/imageability in the twenty-two sub-corpora of the PTRS. This following will describe how this method is implemented in detail. It will accordingly explain how this method works in general and how "seed words" are selected.

Word embeddings

First, the preparation work consists in dividing a corpus into several small sub-corpora according to their historical time span and the transformation of each sub-corpus into a database of word embeddings by FastText (Bojanowski et al. 2017). The currently popular algorithm is the word2vec (Mikolov et al. 2013). This is a program that is especially predictive with regard to learning word embeddings from raw text. However, FastText considers morphological factors and it is suitable for small corpora. This study therefore makes use of FastText to obtain word embeddings that have 300 dimensions. The regular dimension number for word embeddings is usually 100 or 300. 100 dimensions might be a fit for small corpora. However, 300 dimensions might be more helpful for large-scale corpora. The database of word embeddings for this sub-corpus contains the vectors of all the words occurring in this corpus. The word embeddings of this sub-corpus actually contain the information on frequency of each word in this sub-corpus. We then selected a small set of seed words (the number of seed words can be reset according to the corpus size) by considering the factors which will be discussed below.

Implementing dynamic computation

The implementation of this dynamic method is briefly introduced in the following. First, as mentioned above, the *FastText* algorithm was used to derive word embeddings with 300 dimensions from the sub-corpus. Second, an essential component in this method is called the "Sentiment Propagation" ("*SentProp*") (Hamilton et al. 2016a), *SentProp* is used to produce two random walks over a graph of semantic similarities between each word in this sub-corpus (the database of word embeddings of this corpus) and the "seed words". A weighted lexical graph was constructed. In this graph, the nodes are words and edges are constructed by connecting each word to its *n* nearest semantic neighbors as based on the *cosine distance*. Afterwards, the *SentProp* package randomly samples subsets of seed words for each sub-corpus and then it uses word embeddings to calculate the scores with

---

[4] Although Degaetano-Ortlieb and Teich (2018) reported the changes for different types of lexicon and tri-grams by the use of KLD, they did not provide the details and codes on how they implemented the KLD method. After having replicated their study repeatedly, we always failed and therefore attempted to contact them via emails several times. However, they have not responded. This could indicate that their results are unlikely to be replicated. We therefore had to work out several problem-solving strategies. The specific strategies include how to preprocess the corpus, how to remove special characters and symbols (scientific writings are always filled with them), the removal of stop words (the regular way in NLP), as well as how to set up the parameters to select candidate words. Besides these, a "decade" window in our work makes it impossible to carry out comparison with their work using a "two-year" window. We think that the lexicon change could be observed clearly during *two decades* instead of *four years*. The more serious problem is that the data presented (the changes of lexicon KLD) in their study cannot directly support their claim that scientific writings became more professional. Overall, their study cannot be treated as the baseline and the comparisons between their study and the current study are almost impossible.

sampled seed words. Fourth, after the graph has been built, the semantic similarities are transformed into probabilities. We therefore obtain a transition matrix. This matrix provides information on the probability of randomly moving from a word to one of its semantic neighbors [More details can be seen in Hamilton et al. (2016a)]. Two polarity scores are computed in each word and a random subset of seeds: abstraction (a) and concreteness (c). The final score is $c/(c+a)$. The final score is between 0 and 1. When the score of a word approaches 1, this means that the concreteness of this word is high and vice versa. We can obtain the mean concreteness score for each sub-corpus if the sum of the scores of all the words in each sub-corpus is divided by the number of words in this sub-corpus. The mean concreteness score is treated as the degree of concreteness. When the mean concreteness score approaches 0, this means that the concreteness degree is low and vice versa. The great advantage of the method is that it produces a concreteness value that reflects the usage within a specific corpus, i.e. the historical period of use or genre.

A simplified example is used to illustrate our method. Here n is set to 2 and the two nearest semantic neighbors of "sheep" are "dog" and "animal". In the weighted lexical graph, "sheep" will be connected to "dog" and "animal" with an edge weighted by the semantic similarity between them. In the transition matrix, the probabilities of moving from "sheep" to "dog" and "animal" add up to 1, with a higher probability of moving to "dog" because of its stronger semantic connection with "sheep". Now, supposing that two concrete (e.g., "crop" and "cow") and two abstract (e.g., "thought" and "philosophy") seed words are taken. We will not do any random sampling of the seed words here given the small number of seed words used but will perform 100 random walks on the transition matrix from each of the four seed words. One random walk starting from "crop" can be traced as follows: the first step takes us from "crop" to one of its two nearest neighbors (say "farm"), the second step from "farm" to one of its two nearest neighbors (say "animal"), and this procedure was repeated until a specified number of steps (e.g. 1500) have been taken. After the words traversed in the random walk are recorded the next walk is then performed. Among the 100 random walks, the probability of moving from any word to one of its two neighbors will be decided by the transition probabilities in the transition matrix. In general, a random walk from a concrete seed word tend to land on more concrete words than abstract words given that concrete words are more likely to have other concrete walks as their nearest neighbors. The opposite is true for a random walk from an abstract seed word. When 100 walks have been performed from all four seed words, the proportion of walks from the two concrete seed words and the two abstract words landing on "sheep" can be calculated as c (say 0.64) and a (say 0.15), respectively. The concreteness score of "sheep" can then be computed as $c/(c+a)=0.64/(0.64+0.15)=0.81$. The concreteness score of the corpus is the average of the concreteness scores of all words in the corpus.

Word embeddings can represent co-occurrences of each word with other words in a corpus such that each vector describes the typical context of one word. Compared to the first two methods mentioned above, this method which uses word embeddings not only takes word frequency into account but, because of its semi-supervised nature, effectively eliminates the coverage problem by analyzing all words in the corpus, function words included. More importantly, the approach taken by the method is a dynamic one, in that the degree of concreteness of each word is allowed to vary across sub-corpora representing different time periods.

A number of "seed words" that should have a high frequency in each corpus represent the highest and lowest score of concreteness in the corpora that are to be compared here. We use the database of Brysbaert et al. (2014) (40,000 English word lemmas) to obtain the concreteness scores of "seed words" used in our method. The largest

database of imageable subjective norms is the Glasgow database, which contains 5500 words (Scott et al. 2019). Similarly, we use this database of "imageability" to obtain the imageability scores for "seed words". Our method of computing concreteness/imageability is a semi-supervised method: semi-supervised methods require only minimal human supervision (in the form of seed words). In the following, we will discuss how to choose "seed words".

"Seed words"

Snefjella et al. (2019) took fifteen concrete and abstract words as "seed words". They chose seed words that (i) occurred with high frequency in each decade between 1850 and 2000 and were associated with (ii) extreme concrete or abstract ratings. We also applied such criteria to choose seed words. We set up the word frequency (> 200 occurrences) in each decade between 1665 and 1869 and chose words with high concrete or abstract ratings. These "seed words" should be words used frequently in the different sub-corpora. Otherwise, they will cease to be representative of these sub-corpora. In addition to this, we have used the topic modeling (Blei et al. 2003; Hornik and Grün 2011) to train in this corpus in order to obtain "topics". Topic modeling concerns the finding of essential words/terms in a collection of documents that best represents the collection. After that, we obtained a collection of 200 topics (words) which are essential in these scientific papers in this corpus. Moreover, we have considered other factors: (i) these seed words should be covered in a multiple of disciplines; (ii) the concrete "seed words" we have must be found in the collection of topics; otherwise, they will be removed; (iii) meanwhile, we also considered the fact that those words have undergone few changes in meaning. In this sense, these seed words are optimally selected and representative of concreteness, topics and terminologies in diversified disciplines.

The "seed words" should be the same in each sub-corpus that is to be investigated. That is to say, the "seed words" depend on the given number of sub-corpora examined. For example, the number of words shared by all word embeddings transformed by sub-corpora in PTRS is 2149. We then searched for the twenty words of the 980 with highest concreteness scores and for the twenty words with lowest concreteness scores according to the database of concreteness. Meanwhile, the frequency (> 10) and "topics" will be considered in order to remove these unqualified words. The process will continue until forty words with high and low concreteness scores are yielded. The "seed words" of imageability are selected using the same method.

Ultimately, these twenty concrete seed words are *water, body, air, blood, fig, paper, ship, table, moon, foot, glass, tube, telescope, earth, salt, sea, eye, head, ship, paper, sun*. The other group of twenty abstract words are: *hope, enough, belong, extent, justice, manner, theory, value, probable, purpose, contrary, imagine, purpose, believe, certainty, suppose, moment, spirit, circumstances, analogy*. The twenty imageable words in the imageability group are *hand, dog, water, sheep, ear, mountain, horse, nose, coat, eye, table, sky, moon, plant, bird, grass, finger, snow, circle, yellow*. The other group of twenty imageless words are *instance, disposition, philosophy, notion, phenomenon, least, mind, opinion, language, fact, value, chance, process, moment, truth, necessity, origin, theory, advantage, hypothesis*. We use two different sets of
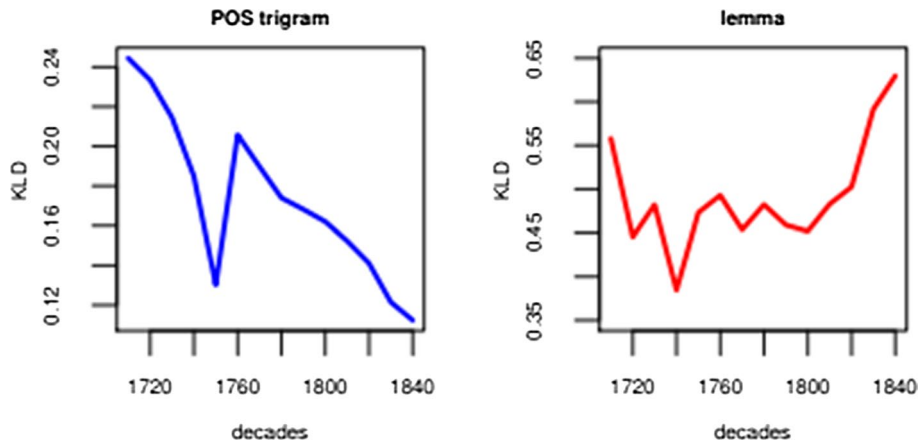
Fig. 2 KLD for POS trigram and lexicon. Lemmas fluctuate in a relatively wild fashion. The KLD of lemmas from 1710 to 1750 exhibited a downward tendency, relatively speaking. Afterwards it remained relatively stable with slight fluctuations between 1760 and 1790. However, after 1790, the KLD began to rise quickly. By contrast, the POS trigram underwent a dramatic drop in spite of some fluctuations

"seed words" derived from the two different databases for two measures in order to guarantee the validity of our method.

Results

Diachronic changes of relative entropy

The first result concerns the changes of lemmas and POS trigrams in the PTRS and it refers to changes of relative entropy. A starting year (1710) and a sliding window (10 years) were selected for this. As the early years (1665–1700) and the final years (1850–1869) might manifest instability and non-typicality with respect to the history of this journal, these two periods were removed. At the lexical level, these candidates of lemma for comparison were selected from those lemmas that were filtered over time. At the grammatical level, those discriminative and significant POS trigrams ranked over time were selected for comparison. The result is shown in Fig. 2.

The right panel in Fig. 2 shows that lemmas fluctuate in a relatively wild fashion, which indicates that the journal had a period of lexical expansion and reduction. In contrast, the left panel in this figure demonstrates that POS trigram undergoes a dramatic drop, indicating that the POS trigrams had a period of consolidation. This tendency becomes clearer in the 19th century. This indicates that grammar might have changed less and became simpler in the PTRS. The evolution of grammatical simplification is quite apparent here. It is widely acknowledged that modern scientific writings were characterized by simple grammar (Gross et al. 2002: 230; Mack 2015). Obviously, the change in the KLD of the POS trigrams accords with widespread expectation about the trend.

We have mentioned the general trends of becoming specialized and professional in scientific discourse. Generally speaking, scientific writings move towards a specialized genre in order to differentiate from other genres. For instance, the massive use of specialized technical terms can strengthen their specialization. Scientific writings need efficient means

of presenting and communicating their findings. For instance, by simplifying grammar in order to aid easy understanding, as well as to partly compensate for the use of specialized terms, scientific discourse became professional. Additionally, scientific writings took their readership into consideration by presenting arguments based on objective analysis rather than merely reporting subjective observations. The following will analyze how our data supports this thesis of a trend towards specialization and professionalization. This evolutionary trend has also partly been confirmed by Degaetano-Ortlieb and Teich (2018). However, the trends of lexicon and POS trigram KLD in our study are partly different from Degaetano-Ortlieb and Teich (2018).

KLD-related measures of probability distributions are also related to the cognitive surprise that readers would experience over time. Cognitive surprise or reader novelty can be used to interpret these changes and these may be helpful in understanding these changes. The continuous decline of the KLD for POS trigrams shows that the degree to which POS trigrams elicit surprise has decreased during the period. In other words, because the POS trigrams have been consolidated and simplified during the period, readers gradually experienced less surprise in encountering them. This means that readers gradually found them easier to process. On the other hand, the KLD of the lexicon fluctuated wildly. From 1710 to 1750, this exhibited a downward tendency, relatively speaking. Afterwards, it remained relatively stable with slight fluctuations between 1760 and 1790. However, after 1790, the KLD began to rise quickly. In the first stage, readers would gradually feel less surprised in encountering the lexicon. The second stage is a little complicated, but in the third stage readers gradually felt surprised in encountering the lexicon. The main reason is that the changes of the KLD for lemmas might have been influenced by how many new words or terms were used or introduced. When fewer new terms were introduced, the KLD exhibited a downward trend, showing that readers did not feel surprised or find the terms difficult. Otherwise, the KLD rose, showing that readers felt surprised or found the terms difficult to process. The data on the changes in the lexicon KLD also demonstrates that specialized and technical terms were used on a large scale in late 19th century and the use of these technical terms reinforced the trend towards specialization in this journal.[5]

Unlike grammar, a dramatic degree of lexical change occurred during the period between the end of 1700s and 1850. The change of the KLD in the lexicon is quite clear during the 19th century. This might indicate that the PTRS largely used specialized technical terms to increase its degree of specialization. The continuous decline of the KLD for POS trigrams, on the other hand, indicates that the PTRS became grammatically simpler, thus allowing for more efficient communication and also increasing its degree of professionalization. The changes in readers' cognitive surprise also reflect the general evolutionary direction. The changes of entropy in the PTRS for POS trigram and lexicon support the thesis that the journal followed the evolutionary trend of gradual increasing professionalization and specialization in authors who are grouped by different decades.

However, the inconsistency between the POS trigram and lexicon might suggest that this tendency did not take place in a straightforward way but that external factors may well have also interfered with this process. In order to gain a deeper understanding of these issues, we now explore this phenomenon with reference to linguistic concreteness.

---

[5] The Jensen–Shannon Divergence (JSD) (Nielsen 2010) is a symmetric and smoothed version of the KL divergence. JSD is another method of measuring the similarity between two probability distribution. We have run all data on the Jensen-Shannon divergence. The results are the same as KLD.
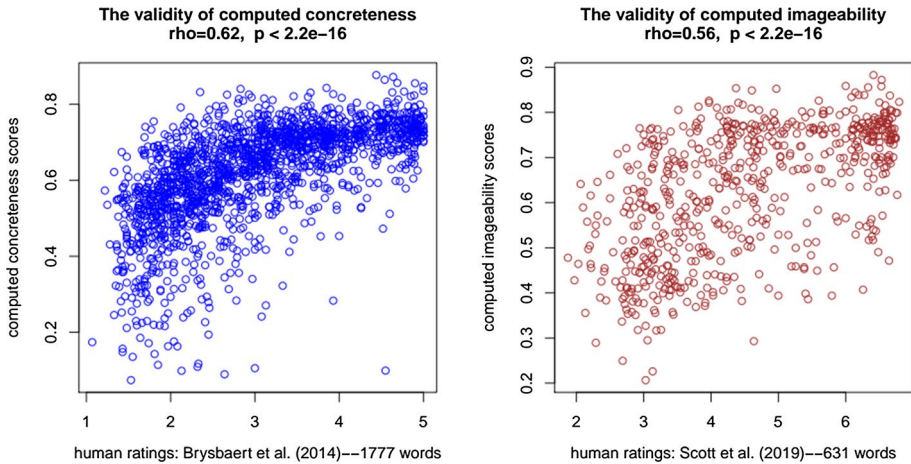
Fig. 3 The correlations between aggregated (1680–1860) computed concreteness/imageability scores and human ratings in concreteness/imageability
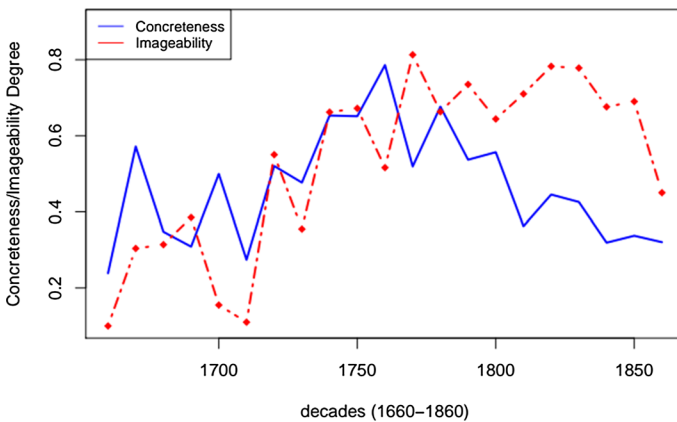

Fig. 4 Linguistic concreteness/imageability of PTRS historically considered. The shape of curve of concreteness and the shape of imageability curve are basically similar. It indicates that the two types of data basically converge. The curves can be divided into *four phases*. The linguistic concreteness/imageability fluctuated stably from 1665 to 1710. The period from the mid-1710s to the mid-1750s saw a gradual increase in the concreteness/imageability, although both were constantly interrupted by fluctuations. The concreteness/imageability remained relatively stable until the beginning of the 19th century. The 19th century shows a drastic drop in concreteness/imageability

Diachronic changes of linguistic concreteness/imageability

First, we need to test the reliability and validity of our computed concreteness and imageability. Figure 3 Specifically, our unsupervised model's concreteness scores correlate with Brysbaert et al. (2014) at $rho = 0.62$ and the imageability scores correlate with Scott et al. (2019) at $rho = 0.56$. The accuracy is similar to Snefjella et al. (2019) ($rho = 0.70$, $p < 0.001$). Overall, the highly significant medium to large correlations (Cohen 1992)

suggested a good level of validity to the scores computed by our dynamic method. This supports the claim of computed scores to reliability and validity.

After using the method of concreteness and imageability, we plot Fig. 4, which represents diachronic changes of linguistic concreteness and imageability respectively in the PTRS. In general, the curve of concreteness is basically consistent with the curve of imageability in each decade. This also proves that the method of using concreteness in detecting language changes is valid and reliable.

Figure 4 shows that the diachronic change of linguistic concreteness is not linear and it does not move towards abstraction in a straightforward way. The diachronic change of linguistic imageability is not linear either and it does not directly tend towards a decrease in imageability. Linguistic concreteness and imageability exhibits stable fluctuations from 1665 to 1710. The period from about the mid-1710s to the mid-1750s saw a gradual increase in concreteness and imageability although both were constantly interrupted by fluctuations. The two measures then remained relatively stable until the beginning of the 19th century. The following decades in the 19th century saw a drastic drop in concreteness. However, imageability had a slightly different status to concreteness during this period because imageability remained stable until 1820s. After the 1820s, it began to decrease rapidly. This is the only noticeable difference between the two curves of concreteness and imageability. However, both curves look similar, forming a parabola-like shape. Generally, the trend towards an increase or decrease in either concreteness or imageability in Fig. 4 would last for an extended period instead of forming a see-saw shape produced by a quick increase or decrease. These changes might have been influenced by external factors. These changes also show that scientists might have used more abstract words in journal writings in 18th century, but after 1800s scientists gradually tended to use more concrete words in journal writings. However, the other social-cultural factors could have influenced this trend. To gain further insight into the causes of these changes, we need to review the important events in the history of the PTRS to see if they are associated with these changes.

Historical events that may have interrupted the evolutionary pattern

Significant historical events in the PTRS

This journal was launched by Henry Oldenburg, the first secretary of the Royal Society in 1665. The early journal relied on Oldenburg's prodigious network of natural-philosophical contacts. After his death, editorial and financial responsibility for the journal was unofficially assumed by his Secretarial successors. After 1713, the journal began to be controlled by Newtonians who were mostly supported by Newton or who highly praised Newton's theories.

The PTRS officially became a Royal Society publication after 1752 and it began to be edited by a committee. A number of highly diverse scientific societies were established on a continually basis after the end of 18 century, such as, the Linnean Society (1788), the Royal Institution (1800), the Horticultural Society (1804) and the Geological Society (1807) amongst others. These newly established specialist scientific societies and institutions began to organize meetings and they launched their own journals. The greater number of scientific journals broke the PTRS's "monopoly" on scientific publications, forcing the PTRS to make great changes, such as expanding its scope and restating its goal. By the 1830s, the journal responded by introducing *expert peer review*. Later in 1886 PTRS split into an 'A' and 'B' series for increased specialization.
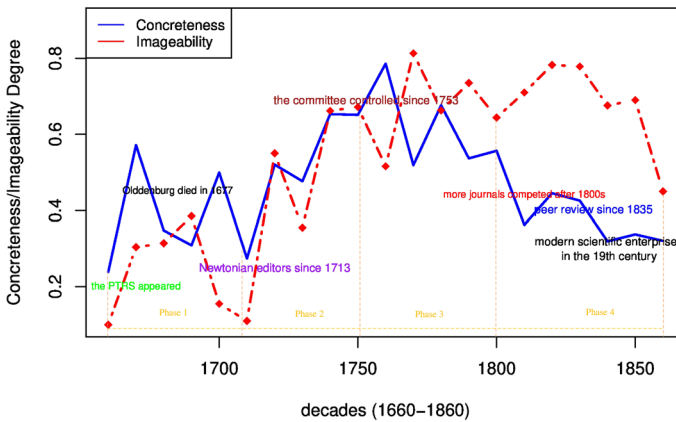
Fig. 5 Concreteness/imageability and historical events. The historical events had an impact on the changes of linguistic concreteness/imageability in the PTRS. These events divide the whole curve of concreteness into four phases, as shown in Fig. 4

We can summarize the history of PTRS during the period from 1665 to 1869 by saying that the significant historical events included changes of editors, publication topics, reviewing policies, and becoming more scientifically professional. These changes will be represented in Fig. 5. It can be assumed that these historical events had great impact on every aspect of this scientific journal, including its language. By comparing Fig. 4 and these events, we found that the changes of linguistic concreteness approximately match the important events in the history of PTRS.

The convergence of KLD and concreteness/imageability coinciding with historical events

By combining the plotted concreteness and the historical events in Fig. 5, we can better see how historical events influenced language use in the PTRS. We also noticed that these historical events were interwoven with the broader socio-cultural environment, thus effecting and influencing language use in scientific discourse.

The specific analysis is contained in the following. After launching the PTRS in 1665, Oldenburg was responsible for editing each issue; after his death, with the exception of the immediate aftermath, which was the notably unstable transition period from 1677 to 1710, the journal remained in continuous publication. Correspondingly, from 1677 to 1710, linguistic concreteness fluctuated mildly, as did linguistic imageability.

In 1713, the editorship was controlled by Newtonians. The journal came out on a quarterly basis for most of this period. In 1718 the Society published a pamphlet that was mainly intended as an advertisement of its accomplishments. This short pamphlet showed which research topics were considered important in this period. This journal not only focused on the issues that Newtonians were interested in, but expanded the number of topics they regarded as significant. These new topics mostly concerned accurate accounts of all uncommon appearances in natural phenomena, new discoveries in natural history as well as new experiments. The papers concerning novelties and innovations were usually constituted by plain descriptions rather than highly technical writings (more details in Atkinson 1998: 22–24). The selection of papers was also influenced by the changes of

editors. The ideological shift in the journal contributed to linguistic shifts. Under the influence of editorial policy and the scientific environment, the scientific papers at this stage preferred plain language and a "descriptive" style. This could have brought about a relatively rapid increase in both concreteness and imageability during the period (1713–1753), but with some fluctuations. By contrast, from Fig. 2, we found that the KLD of lemmas from 1710 to 1750 exhibited a relatively downward tendency. Due to the frequent use of common words, the KLD of lemmas declined quickly during the period.

The journal was controlled by the Committee of Papers of the Royal Society rather than the secretary after 1753. The committee mostly based its judgments about the publication and rejection of papers based on the short abstracts read during its weekly meetings. This policy made the choice of topics and style preferences relatively consistent and fair, in contrast to decisions made on the basis of the preferences of an individual (editor). The implementation of a review panel by the journal may have led to stability in concreteness during the period between 1760 and 1800, albeit with some fluctuations. The similar stability in imageability with fluctuations can also be seen over the same period. Unlike the previous stage, very few new findings were marked for publication, so the concreteness hardly increased and remained at a fairly stable level. Interestingly, the KLD of the lemma performed similarly to the concreteness/imageability. It remained relatively stable with some fluctuations between 1760 and 1790

The early journals published papers that can be called "descriptive". A scientist would typically report that "First, I saw this, and then I saw that" or "First, I did this, and then I did that" (Gross et al. 2002). The articles often intended to establish the credibility of these eye-witness accounts. This descriptive style was appropriate to the kind of science that was then reported, as particularly shown in these papers reported in the PTRS under the editorship of the Newtonians (1713–1752). However, the language, style and structure of scientific writings changed massively in the 19th century.

At the start of the 19th century, other journals from other societies were created, competing with the PTRS. Other journals began to specialize in their own fields and the papers in these journals were fairly professional and scientific. In this situation, the PTRS had to adjust through series of strategic changes, including the use of the peer review. The policy of peer review was taken up in 1835 and the journal settled on a very clear goal: that of making the journal abstract. These strategies were quite effective. They made the journal more scientific, and thus, more competitive in the field of science journals. The other effect of these strategies was that of making the language in papers more professional and technical. The drop in linguistic concreteness, which conformed with the journal's goal, was likely the result of these strategies. However, this effected imageability after 1820s, i.e. 20 years later than the changes in concreteness. The inconsistency between concreteness and imageability might be caused by the different "seed words" selected in our method.

It is commonly acknowledged that by the 19th century, science was beginning to move fast and in increasingly sophisticated ways and that the practice of science became more professionalized and institutionalized in a manner that continued throughout the 20th century. According to the study of Cahan (2003), during the 19th century, much of the modern scientific enterprise took shape: scientific disciplines were formed and institutions and communities were founded. Many of these organizations began to publish journals and proceedings. These journals started specializing in the different disciplines, as detailed above. The scientific writings became a presentation of an argument rather than a simple report or description. The professionalization and specialization of science forced scientific discourse to become professional and specialized in its language and structure, for example, using technical terms rather than plain words and having a structured argument.

These changes had great impact on language in scientific writings. The data show that the following decades in the 19th century therefore exhibited a drastic drop in the concreteness and imageability in the PTRS. On the contrary, the KLD underwent a rapid rise during the period due to the frequent use of specialized and technical expressions. These changes are described quite well by Mack (2015).

> Early on there was little technical vocabulary, and scientific literature could be read by any educated person. The growth of science led to specialization, which led to the invention of a specialized vocabulary. In the 18th and 19th centuries… the language of early science articles was full of complicated clauses and long sentences….Today, scientific language is characterized by simple sentence structures and simple verbs… full of technical terms only understood by similarly informed readers. (Mack 2015)

As mentioned above, the rapid increase of the KLD in the lexicon during the 19th century reflects the dramatic change in lexical uses that occurred during the period between the end of 18th century and 1850. Meanwhile, linguistic concreteness underwent a rapid decline during the same period. Similarly, linguistic imageability also underwent a rapid decline over this period although it was stable in the initial 20 years of this period. The performance of the KLD of the lemma and concreteness shows that the PTRS largely used specialized technical terms in order to become increasingly specialized. The continuous decline of KLD for POS trigrams, on the other hand, suggests that the PTRS simplified grammar to enable efficient communication, helping to realize its professionalization.

To sum up, when Figs. 2 and 5 were combined, it was clear that the two data groups were well mapped. The changes of KLD for the lexicon are almost inversely consistent to the changes of linguistic concreteness/imageability. The right panel of Fig. 2 (lexicon KLD) looks like a parabola(u-shaped). By contrast, Fig. 4 (concreteness/imageability) looks like an anti-parabola (n-shaped). Basically, the shapes in two figures have two reversed orientations. This indicates that the data from two computational methods can be well mapped (converged) to support the argument that this journal followed the evolutionary trend of becoming more professionalized and specialized. But it also shows that the journal was greatly influenced by historical events associated with broader social-cultural changes. Although the POS trigram cannot represent all grammatical constructions, it is at least a fairly typical and largely grammatical structure. The KLD of POS trigrams had been decreasing during the period of the 18th and 19th centuries. This indicates that the grammar was becoming simpler. This accords the general expectation described by Mack (2015). We therefore think that the results analyzed here are very helpful in understanding how scientific culture was interwoven with human cognition and culture.

Discussion

In the introductory section, we put forward the hypothesis that the evolutionary trend in science writings ought to be relatively linear and that slight fluctuations therein would be caused by socio-cultural factors. However, the real evolutionary route is not linear. This evolutionary route was interrupted by external factors far more than we had initially expected. This meant we had underestimated the role of external factors. This also shows that external socio-cultural factors exerted an immeasurable impact on the developmental process of scientific discourse. However, and in accordance with our initial hypothesis, the overall evolutionary trend is ultimately one in which the language in PTRS became

increasingly abstract and less imageable, thus conforming to our initial hypothesis. That is the main finding of the present study.

Our findings are partly consistent with the previous studies. For instance, Banks (2008) used 30 research articles from the PTRS (1700–1980) to investigate their development. He found that the frequency of nominalization increased in the period in question. The frequent use of nominalization indicates that the word abstractness increased. However, the number of research articles was rather small and these thirty articles were not evenly distributed over the 280 year period. This makes it difficult for the study to detect the changes in nominalization in each decade. In addition to this, the PTRS was split into an 'A' and 'B' series in 1886 and afterwards into still further series again. This indicates that after 1880s the PTRS should not be treated as being the same as that which was published under this name prior to the 1880s. Biber and Gray (2016) investigated changes of linguistic styles in academic writing from the 18th century to the present. Biber and Gray (2016) also found an increase of frequency in nominalization and common nouns. The increase in nominalization and common nouns supports the thesis that scientific writings became ever more abstract. Meanwhile, Biber and Gray (2016) emphasize changes in phrases. They found that the following aspects of phrases increased in frequency: noun phrase pre-modifiers, specific phrase prepositions and phrasal devices modifying head nouns. They concluded that the grammatical style of academic research writing was characterized by a compression of structure (the compression of phrase structures). Their finding is consistent with our finding concerning the KLD of POS trigrams, namely that POS trigrams were consolidated and simplified during the period shown in the left panel of Fig. 2. Our finding has also partly been confirmed by Degaetano-Ortlieb and Teich (2018). In addition, the trend of using more abstract terms in scientific writings seems to continue after 1870s. After analyzing abstracts published between 1881 and 2015 from 123 scientific journals, Plavén-Sigray et al. (2017) found there might have been a growing use of general scientific terms. This trend can be confirmed as occurring after the 1800s and as continuing after the 1900s.

Although the two types of methods represent different dimensions, both groups of data can help cross-verify the evolutionary pattern of language in a scientific genre and address the questions raised in the Introduction. As shown in piecewise-linear sections in the plots (Figs. 2, 5) of the two measures in KLD and linguistic concreteness/imageability, they coincide with historical periods that would influence the language produced in the PTRS. The consistency between concreteness and imageability shows that the method of diachronic concreteness is reliable and effective. The trends described in these plots purportedly support a narrative of increasing professionalization and specialization in science over the first two hundred years of the PTRS. All these clearly demonstrate how external factors influenced language use in the PTRS and how language in PTRS changed to become more professionalized and abstract in the 19th century.

After some long-term and unexpected fluctuations, scientific writings did become more abstract and less imageable in the 19th century. To be exact, after 1800s, a continuous and rapid increase of the KLD for the lexicon also indicates that authors introduced more technical terms in their writings. By contrast, authors tended to use simpler grammar in the PTRS. Clearly, two groups of data are matched to support our revised argument which is based on the initial hypothesis we proposed in the Introduction. Although the PTRS just is one example of scientific writings overall, its changes should still accord with the general evolutionary pattern of the scientific world. However, this pattern was violently interrupted by the external factors that were mentioned above. These factors mainly involve three domains: the journal itself, the overall development of science, as well as the broader social-cultural changes analyzed above. Factors in

the three domains interacted with each other to influence the evolution of scientific discourse. Very interestingly, the degree of abstractness/imageability (or lexicon KLD) of the PTRS in 19th century returned to the same level it was at in the 1670s. This process is clearly interrupted by the various factors we have discussed. The finding seems to support the idea of Gross et al. (2002: 138) "Stylistically, 19th-century scientific prose appears to more closely resemble its 17th century origins than the highly compressed, neutral, monotonal prose of the late 20th century." However, it does not mean that scientific writings return to the status of writings in 17th century. But this is an issue that deserves to be researched more closely.

The evolutionary pattern in scientific writings did not take place in a straightforward way, but rather it fluctuated wildly. These fluctuations were clearly influenced by the events from the aforementioned three domains. The two groups of data (the KLD in the lexicon and linguistic concreteness/imageability) are well mapped with historical events in this journal. Specifically, scientific writings did not experience a direct movement towards becoming a professionalized and specific genre. Contrary to the widespread belief, the evolutionary direction of language in this journal is not necessarily a linear one. In fact, the direction of linguistic evolution in this journal was greatly influenced by external social-cultural factors. Generally, the diachronic changes in relative entropy and linguistic concreteness/imageability in the PTRS actually help in revealing the overall development of science in the last 300 years.

This novel quantitative method is capable of examining the diachronic changes in the language of scientific discourse. By analyzing full-text documents in the PTRS, we find that language in the PTRS evolved towards an increasing professionalization and specialization in the authors who are grouped together by decades. In fact, this quantitative method is effective in analyzing full texts. This method is as useful as linguistic complexity in analyzing scientific writings (Lu et al. 2019a, b; Chen et al. 2020). We believe that this quantitative method will reveal yet more scientometrics findings when the bibliometric data becomes available and can make a great contribution to full-text analysis. For instance, relative entropy and linguistic concreteness can be applied to investigating how either of the measures (or both) varies with the changes of scientific impacts or authors with different background.

However, some limitations in the current study should be noted. First, the current findings are based on "the limited sample". That is to say, the size of the PTRS (1665–1869) is a bit small for scientific writings in journals over the history. The limited availability of more scientific writings might constraint quantitative studies. The second limitation is that we did not include the other types of scientific writings (books, reports, etc.) during this period. It would perhaps be interesting to look at samples from other kinds of scientific writings, and make comparisons between the PTRS and the other types of scientific writings. Third, after the PTRS split into series "A" and "B" in 1880s, more changes later are fantastic steps of scientific progress. It should be amazing to make a diachronic study of the PTRS after 1880s. However, the current study only considered the PTRS published before 1869. Forth, some traditional bibliometric factors (e.g., keyword frequency/ratio, bibliographic coupling, co-word analysis, citation) could be included, and this could show the significance of this research. However, without databases containing bibliometric factors on the PTRS, these bibliometric studies would be difficult to implement. We can make further studies by considering these factors, such as changes of topic words, co-authors etc., to such investigations in scientific writings published in modern journals by using the methods proposed in the current study.

Concluding remarks

By combining the two computational approaches (relative entropy and word-embedding concreteness/imageability), this study created a novel quantitative method for examining the diachronic changes in the language of scientific discourse (the PTRS) over two hundred years. Because the two methods are mutually complementary, an integrated approach can explore diachronic linguistic changes in the PTRS in a comprehensive manner. The diachronic trend in scientific discourse can be represented in linguistic data and therefore can be detected through measuring the changes of entropy and linguistic concreteness. Our research thus creates a novel quantitative methodology and applies this to the examination of diachronic changes in language. This is the main contribution of the present study.

After the whole *Royal Society Corpus* was divided into twenty-two sub-corpora according to one decade, the KLD measures were used to detect the diachronic changes of POS trigram and lemma among the twenty-two sub-corpora, as well as to quantify the changes in cognitive surprise between the two decades in order to estimate whether readers encountered more/less difficulties in processing the language used. Linguistic concreteness and imageability based on word embeddings produced by the word embedding algorithm were also used to quantify the diachronic changes of word concreteness and imageability over the twenty decades. The data from the two separate methods were consistent. Based on the analysis of the data, we found that the evolutionary direction was not necessarily linear but rather that it was interrupted by social and cultural factors. Linguistic concreteness/imageability and the KLD of lexicon changed in the light of such events. In contrast, the simplification of grammar continued almost uninterrupted. The data from both methods are consistent in supporting the thesis that this journal followed the evolutionary trend of becoming increasingly professionalized and specialized, but that this process was also greatly influenced by historical events and broader social-cultural factors. This finding is distinct from the initial hypothesis that scientific discourse in journals underwent a linear evolutionary process of increasing professionalization and specialization.

Overall, this study clearly demonstrated how the language of scientific writings in early journals followed an inherent evolutionary pattern and how its evolutionary route was greatly influenced by external factors. The current study uncovered the general developments in science writing over the last 300 years and it thus helps in understanding how scientific culture was interwoven with human cognition and culture. This study provides a perspective to overview language style of scientific writings in early journals (i.e., in 17th, 18th and early 19th centuries). The quantitative methods proposed by the current study provide a new perspective in full-text analysis.[6] However, this study has the limitations of sample size and bibliographic factors. In future, we can make further scientometrics studies of further factors. We believe that further interesting findings can be yielded with respect to these other variables.

---

[6] All programming codes (Perl, Linux shell and R) and data used in this study are available in https://osf.io/n9cj5.

References

Atkinson, D. (1998). *Scientific discourse in sociohistorical context: The Philosophical Transactions of the Royal Society of London, 1675–1975*. London: Routledge.

Bakarov, A. (2018). *A survey of word embeddings evaluation methods*. arXiv preprint arXiv:1801.09536.

Banks, D. (2008). *The development of scientific writing. Linguistic features and historical context*. London: Equinox.

Barron, A. T., Huang, J., Spang, R. L., & DeDeo, S. (2018). Individuals, institutions, and innovation in the debates of the French revolution. *Proceedings of the National Academy of Sciences, 115*(18), 4607–4612.

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic english: Linguistic change in writing*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written english* (Vol. 2). London: Longman.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics, 5,* 135–146.

Boyack, K. W., Small, H., & Klavans, R. (2013). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology, 64*(9), 1759–1767.

Boyack, K. W., van Eck, N. J., Colavizza, G., & Waltman, L. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. *Journal of Informetrics*, *12*(1), 59–73.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods, 46*(3), 904–911.

Cahan, D. (2003). *From natural philosophy to the sciences: Writing the history of nineteenth-century science*. Chicago: University of Chicago Press.

Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science, 41*(4), 892–923.

Casadevall, A., & Fang, F. C. (2014). Specialized science. *Infection and Immunity*, 1355–1360.

Chen, B., Deng, D., Zhong, Z., & Zhang, C. (2020). Exploring linguistic characteristics of highly browsed and downloaded academic articles. *Scientometrics, 122*(3), 1769–1790.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Dawson, G., Lightman, B., Shuttleworth, S., & Topham, J. R. (2020). *Science periodicals in nineteenth-century Britain: Constructing scientific communities*. Chicago: University of Chicago Press.

De Saussure, F. (2011). *Course in general linguistics*. Columbia: Columbia University Press.

Degaetano-Ortlieb, S., & 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* (pp. 22–33).

Ding, Y., Song, M., Han, J., Yu, Q., Yan, E., Lin, L., et al. (2013). Entitymetrics: Measuring the impact of entities. *PLoS ONE, 8*(8), e71416.

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology, 65*(9), 1820–1833.

Elgendi, M. (2019). Characteristics of a highly cited article: A machine learning perspective. *IEEE Access, 7,* 87977–87986.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences, 115*(16), 3635–3644.

Glenisson, P., Glänzel, W., & Persson, O. (2005). Combining full-text analysis and bibliometric indicators. A pilot study. *Scientometrics, 63*(1), 163–180.

Gould, S. J. (2002). *The structure of evolutionary theory*. Harvard: Harvard University Press.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 193–202.

Gross, A. G., Harmon, J. E., Reidy, M., Reidy, M. S., et al. (2002). *Communicating science: The scientific article from the 17th century to the present*. Oxford: Oxford University Press.

Hamilton, W. L., Clark, K., Leskovec, J., & Jurafsky, D. (2016a). Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the conference on empirical methods in natural language processing* (p. 595).

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016b). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th annual meeting of the association for computational linguistics* (Volume 1: Long Papers) (pp. 1489–1501).

Harmon, J. E. (1989). The structure of scientific and engineering papers: A historical perspective. *IEEE Transactions on Professional Communication, 32*(3), 132–138.

Hill, F., Korhonen, A., & Bentz, C. (2014). A quantitative empirical analysis of the abstract/concrete distinction. *Cognitive Science, 38*(1), 162–177.

Hills, T. T., & Adelman, J. S. (2015). Recent evolution of learnability in American English from 1800 to 2000. *Cognition, 143,* 87–92.

Hornik, K., & Grün, B. (2011). TopicModels: An r package for fitting topic models. *Journal of Statistical Software, 40*(13), 1–30.

Houghton, B. (1975). *Scientific periodicals: Their historical development, characteristics and control*. Bingley: Shoe String Press Inc.

Hutchison, J., Cunningham, S. J., Slessor, G., Urquhart, J., Smith, K., & Martin, D. (2018). Context and perceptual salience influence the formation of novel stereotypes via cumulative cultural evolution. *Cognitive Science, 42,* 186–212.

Iliev, R., Hoover, J., Dehghani, M., & Axelrod, R. (2016). Linguistic positivity in historical texts reflects dynamic environmental and psychological factors. *Proceedings of the National Academy of Sciences, 113*(49), 7871–7879.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295–1306.

Jacob, M. C. (1988). *The cultural meaning of the scientific revolution*. Philadelphia: Temple University Press.

Jurafsky, D., & Martin, J. (2008). *Speech and language processing*. Upper Saddle River, NJ: Pearson Prentice Hall.

Kermes, H., Degaetano-Ortlieb, S., Khamis, A., Knappen, J., & Teich, E. (2016). The royal society corpus: From uncharted data to corpus. In *LREC*.

Klingenstein, S., Hitchcock, T., & DeDeo, S. (2014). The civilizing process in London's Old Bailey. *Proceedings of the National Academy of Sciences, 111*(26), 9419–9424.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79–86.

Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. (2018). *Diachronic word embeddings and semantic shifts: A survey*. arXiv preprint arXiv:1806.03537.

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (taales): Version 2.0. *Behavior Research Methods, 50*(3), 1030–1046.

Lindsey, D. (1978). *The scientific publication system in social science*. San Francisco: Jossey-Bass Inc Pub.

Liu, S., & Chen, C. (2013). The differences between latent topics in abstracts and citation contexts of citing papers. *Journal of the American Society for Information Science and Technology, 64*(3), 627–639.

Lu, C., Bu, Y., Dong, X., Wang, J., Ding, Y., Larivière, V., et al. (2019a). Analyzing linguistic complexity and scientific impact. *Journal of Informetrics, 13*(3), 817–829.

Lu, C., Bu, Y., Wang, J., Ding, Y., Torvik, V., Schnaars, M., et al. (2019b). Examining scientific writing styles from the perspective of linguistic complexity. *Journal of the Association for Information Science and Technology, 70*(5), 462–475.

Lu, C., Ding, Y., & Zhang, C. (2017). Understanding the impact change of a highly cited article: A content-based citation analysis. *Scientometrics, 112*(2), 927–945.

Mack, C. A. (2015). 350 years of scientific journals. *Journal of micro/nano-lithography, MEMS, and MOEMS, 14*(010101), 1–010101.

Marks, J. (1986). *Science and the making of the modern world*. Kern: ERIC.

Martin, J. R., & Veel, R. (2005). *Reading science: Critical and functional perspectives on discourses of science*. London: Routledge.

Mckeown, K., Daume, H., Chaturvedi, S., Paparrizos, J., Thadani, K., Barrio, P., et al. (2016). Predicting the impact of scientific concepts using full-text features. *Journal of the Association for Information Science and Technology, 67*(11), 2684–2696.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176–182.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv preprint arXiv:1301.3781.

Morris, P., Hart, J. C., Henderson, L., & Sayer, P. (2013). *Milestones of science and technology: Making the modern world*. London: KWS PUBL.

Murdock, J., Allen, C., & DeDeo, S. (2017). Exploration and exploitation of Victorian science in Darwin's reading notebooks. *Cognition, 159,* 117–126.

Nielsen, F. (2010). A family of statistical symmetric divergences based on Jensen's inequality. arXiv preprint arXiv:1009.4004.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1p2), 1.

Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *Elife, 6,* e27725.

Sayood, K. (2018). Information theory and cognition: A review. *Entropy, 20*(9), 706.

Schmid, H. (1999). Improvements in part-of-speech tagging with an application to German. In *Natural language processing using very large corpora* (pp. 13–25). Berlin: Springer.

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow norms: Ratings of 5500 words on nine scales. *Behavior Research Methods, 51*(3), 1258–1270.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.

Shennan, S. (2009). *Pattern and process in cultural evolution*. Berkeley: University of California Press.

Sherwin, W. B. (2018). Entropy, or information, unifies ecology and evolution and beyond. *Entropy, 20*(10), 727–741.

Snefjella, B., Généreux, M., & Kuperman, V. (2018). Historical evolution of concrete and abstract language revisited. *Behavior Research Methods, 51*(4), 1693–1705.

Spier, R. (2002). The history of the peer-review process. *Trends in Biotechnology, 20*(8), 357–358.

Ure, J. (1982). Introduction: Approaches to the study of register range. *International Journal of the Sociology of Language, 1982*(35), 5–24.

Affiliations

Kun Sun[1,2] · **Haitao Liu[2,3]** · **Wenxin Xiong[4]**

[1] Department of Linguistics, University of Tübingen, Tübingen, Germany

[2] Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China

[3] Department of Linguistics, Zhejiang University, Hangzhou, China

[4] School of International Chinese Studies, Beijing Foreign Studies University, Beijing, China