

Secondary Publication



Watson, Donald; Krug, Manfred; Carbon, Claus-Christian

From a one-dimensional hierarchy to a multidimensional space of scientific english : the relationship between a discipline's place in the hierarchy of the sciences and its use of grammatical structures

Date of secondary publication: 05.10.2026

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-117555x

Primary publication

Watson, Donald; Krug, Manfred; Carbon, Claus-Christian (2026): From a one-dimensional hierarchy to a multidimensional space of scientific english, in: Scientometrics : an international journal for all quantitative aspects of the science of science, communication in science and science policy, Dordrecht [u.a.]: Springer Science + Business Media B.V., Vol. 131, No. 10, pp. 7227–7249, doi: 10.1007/s11192-026-05780-1.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>



From a one-dimensional hierarchy to a multidimensional space of scientific english: the relationship between a discipline's place in the hierarchy of the sciences and its use of grammatical structures

Don Watson¹ · Manfred Krug¹ · Claus-Christian Carbon²

Received: 8 February 2026 / Accepted: 7 August 2026 / Published online: 5 September 2026
© The Author(s) 2026

Abstract

The idea of a Hierarchy of the Sciences (HoS), first proposed by Auguste Comte in the 19th Century, argues that scientific disciplines can be ordered in a hierarchy of complexity, with the inorganic sciences (Astronomy, Physics) being the least complex and the organic sciences (Biology, Sociology) being the most complex. Nearly 200 years later, this order still reflects widely held attitudes towards the relationships between and within academic (sub-) disciplines. HoS-related hypotheses have been supported by a variety of studies using bibliographic, linguistic and scientometric data. The present study investigates whether there is further linguistic, more precisely, grammatical evidence to support or qualify the HoS. We collected a corpus of 39,809 research article abstracts (7,630,993 words) from the Web of Science (WoS) database, performed a multidimensional analysis to identify typical patterns of co-occurring grammatical features and then compared the results of the sub-corpora of the disciplines commonly included in research on the HoS (Arts & Humanities, Sociology, Psychology, Biology, Chemistry, Physics, Astronomy & Astrophysics and Mathematics). Although some of our findings replicated those of previous studies, notably that the grammatically most similar disciplines are those adjacent in traditional arrangements on the HoS, we did not find evidence of a simple one-dimensional hierarchy. Instead, our findings indicate that typical grammatical patterns in the disciplines at the extremes of the traditional hierarchy also resemble one another. This suggests that, as far as grammatical patterns are concerned, representing disciplines in a two-dimensional space of the sciences is superior to a one-dimensional hierarchy.

Keywords Specificity · Discourse community · Multidimensional analysis · Hierarchy of the sciences · Research article · Communication · Disciplines · Scientometrics

✉ Don Watson
don.watson@uni-bamberg.de

¹ Department of English Linguistics, University of Bamberg, Bavaria, Germany

² Department of General Psychology and Methodology, University of Bamberg, Bavaria, Germany

Introduction

This example from the popular internet webcomic *xkcd*¹ (Monroe, 2008) illustrates a commonly held perception of the relationships between scientific disciplines. This is often described in terms of a hierarchy that places natural sciences such as physics near the top and social sciences such as sociology near the bottom (Cole, 1983; Fanelli, 2010). This perception is also often expressed in terms of the ‘hardness’ or ‘softness’ of an academic discipline, with social sciences being described as soft and physical sciences as hard. This tendency can be traced back at least to Auguste Comte in the 19th Century, who ranked scientific disciplines according to a so-called hierarchy of the sciences (HoS). Although this hierarchy categorises academic disciplines primarily by the object of their study, it is also widely held that the hierarchy reflects other aspects such as methodology, discourse conventions and scientific rigour that are related to the object of study. To go back to Comte’s (1853) text:

Thus, we have before us five fundamental Sciences in successive dependence, — Astronomy, Physics, Chemistry, Physiology, and finally Social Physics.² The first considers the most general, simple, abstract, and remote phenomena known to us, and those which affect all others without being affected by them. The last considers the most particular, compound, concrete phenomena, and those which are the most interesting to Man. (pp 29-30)

In this description, the criterion for the ranking is simple and abstract at one extreme and particular and concrete at the other. Further, Comte divided the sciences into two categories, organic and inorganic. This would seem to be the beginning of the ‘soft’ (organic) and ‘hard’ (inorganic) distinction. Although not explicitly stated, it would seem from this that “up” is the opposite to that described by Cole, Fanelli and *xkcd*. Simonton (2015) makes a similar point about the arbitrariness of what is the “top” of the hierarchy while noting that, “Comte himself viewed sociology as the last and greatest of the sciences.” Unsurprisingly, it is in a paper by Haslam and Baes (2023) investigating possible bias in recognition by learned academies in Australia that the order is described as, “moving up the hierarchy from mathematics to sociology.” In the case of *xkcd* then, “purity” is simply a re-framing of Comte’s “general, simple (and) abstract.” The comic is, of course, a joke, and shouldn’t be taken too seriously. But the joke works because it reflects widely held opinions, including, at least in part, scientists’ perceptions of their own disciplines.

Expanding the hierarchy

The list of disciplines included in Comte’s hierarchy is indicative of its age. In the more than 170 years since, many new scientific disciplines have emerged and several studies have attempted to expand the HoS by identifying the place of other disciplines within the hierarchy. Smith et al. (2000) looked at the use of graphs in seven disciplines (Physics, Chemistry, Biology, Medicine, Psychology, Economics and Sociology) and ten sub-disciplines of

¹ Thanks to Randall Monroe of *xkcd.com* for his generous attitude towards copyright.

² The term Social Physics refers to Sociology. “Sociology (which is the term I may be allowed to invent to designate Social Physics)” (p.444, Comte, 1853).

Psychology to assess the relative hardness of these fields. They found that the proportion of the total page area of research articles (RAs) dedicated to graphical data displays correlated closely with the perceived hardness of the academic disciplines studied. Simonton (2015) evaluated the HoS concept using measures explicitly intended to gauge the relative scientific status of various disciplines in order to find the place of Psychology within the hierarchy. Measures included the ratio of theories to laws in textbooks, consultation rate, graph prominence, peer evaluation consensus, early impact rate, citation concentration and obsolescence rate. The results supported the HoS order with Psychology between Biology and Sociology but significantly closer to Biology. Hudson (2017) attempted to find Economics' place as either "a science or a social science" based on bibliometric data, including title length and the use of colons or question marks. Using cluster analysis, Hudson was able to identify clusters that closely resembled the natural sciences, social sciences, and arts and humanities. Benjafield (2020) looked explicitly at the classic HoS disciplines, comparing the vocabulary employed by each. The results were generally consistent with the HoS, with each subject being most similar to its neighbour on the hierarchy. He then expanded the disciplines to include Mathematics, Computing, Economics, Political Science, Philosophy and Linguistics. Here he found interesting similarities between the vocabulary used in Mathematics, Philosophy and Linguistics. These are disciplines, especially Mathematics and Philosophy, which would usually be considered at opposite ends of a hard/soft continuum. This suggests again that as soon as more disciplines than those in Comte's original classification are included, the hierarchy becomes less clear.

Ordering the hierarchy

Beyond simply a vague notion of a hard-to-soft continuum, some thought has been given to how disciplines might fit on this continuum and in what order. In 1973, Biglan assessed scholars' judgements of the similarities between disciplines and identified three dimensions upon which these judgements were based. These were the existence of a paradigm, concern with application and life systems. The first of these was broadly similar to the HoS and the hard/soft differentiation. More recently, the difference between 'hard' and 'soft' science is summarised by Fanelli (2010) as the data and theories "speak more for themselves" in 'hard' science, whereas other psychological and social factors of the researchers play a greater role in the 'softer' sciences. Fanelli's study analysed 2,434 research articles (RAs) to determine how many of these papers reported positive or negative support for a tested hypothesis. He hypothesised that if the HoS is valid, then research in 'softer' sciences should face fewer "constraints to their conscious and unconscious biases" (Fanelli, 2010, p 1) and thus report more positive results. The study found that while the nature of tested hypotheses and the methodological rigour used vary systematically in accordance with the HoS, the social sciences are far from completely subjective and differ from the natural sciences only "by a matter of degree" (Fanelli, 2010, p 5).

Li et al. (2025) offer an alternative hypothesis for the difference between hard and soft sciences in a study that used a large language model to identify the use of lexical metaphors in RA abstracts. They found that lexical metaphors were used more frequently in softer sciences (Medical/Life Sciences and Social Sciences) than in the harder sciences (Applied Sciences, Sciences and Multi-disciplinary Sciences). They first considered more traditional explanations for the preference for metaphors in the softer sciences and attributed the differences to researchers from the softer sciences being more motivated to "clearly explain

their research purposes and arguments,” whereas researchers from hard disciplines, “primarily focus on explaining their methods and results.” (p.10) However they concluded that the differences are more likely due to the following factors: “the academic backgrounds of readers in soft disciplines are more diverse, whereas those in hard disciplines are more specialised and concentrated,” and, “the professional knowledge possessed by readers in the soft disciplines may be more fragmented, while the accumulation of knowledge of readers in the hard disciplines follows a more linear trajectory, and more consensus exists for readers regarding their professional knowledge.” (p 10).

Fanelli and Glänzel (2013) also explored this idea that the harder sciences are more able to achieve consensus due to their, relative to the social sciences, lack of complexity. In fact, they consider this the, “fundamental prediction made by a modern version of the HoS,” (p 2) despite the fact that the concept of complexity is itself the matter of much debate (Zuchowski, 2012). In this study, Fanelli and Glänzel (2013) compared a variety of bibliometric measures hypothesised to indicate scholarly consensus across 12 academic disciplines. The parameters included the number of authors, length of article, number of cited references, proportion of books in references, age of references, diversity of cited sources, relative title length, use of first person in abstracts and likelihood to share references with other papers in the sample. Of these parameters, the use of first-person pronouns was exceptional in that it is purely linguistic rather than bibliometric in nature. The authors reasoned that harder sciences aim to make universal claims, so the writing was expected to be more impersonal. Thus, they hypothesised that writing in disciplines identified as harder would use fewer first-person pronouns. However, apart from the number of authors of research articles, this linguistic parameter was the only parameter that did not conform to their expectations regarding the HoS. Despite the lack of empirical support in this study, it does indicate an expectation that discipline hardness should be reflected in the language used by authors.

The hierarchy and language

The idea that different academic disciplines’ knowledge-making activities will be reflected in their linguistic choices is a foundation of the concept of genre, especially as it relates to academic communication. As Hyland (2001) stated, “writers’ decisions are closely related to the social and epistemological practices of their disciplines” (p. 224). In this study, Hyland (2001) examined self-mention across eight fields. These were divided into soft (Marketing, Philosophy, Applied Linguistics and Sociology) and hard fields (Biology, Physics, Electronic Engineering and Mechanical Engineering). He found, unlike Fanelli and Glänzel (2013), that soft fields used first-person pronouns three times more often (per paper) than hard fields. Thus, we have an initial connection between the hardness or softness of an academic discipline and the grammatical choices made by its authors.

Abstracts

Arguably, the best place to look for authors to exhibit their disciplinary bona fides in language is in RA abstracts. Abstracts are a vital part of any RA as they summarise and showcase the content of the whole article. Indeed, their role is only increasing as many people search for RAs in online databases that contain the abstracts but not the whole text.

Therefore, after having one's attention caught by the title, it is the abstract that convinces the reader to download the full text and read on.

Although the abstract mirrors the full text, research has shown that while abstracts contain some combination of background, aim, methods, results and conclusion moves (Hyland, 2000) similar to the sections of typical RAs, they also exhibit some important differences. Valipour et al. (2016), for instance, have shown that abstracts contain more nouns, lexical density, and longer sentences, and the rest of the text contains fewer modals and adverbs. Both supporting and elaborating these findings was Holtz (2011), who found that proper nouns occur twice as frequently in abstracts, modals occur half as frequently, adverbs and present tense occur less frequently, while past participles occur slightly more often. So, although the language in abstracts cannot be considered entirely representative of that of the rest of the article, the language used in them is possibly more important than that used in the rest of the text.

Fortunately, an informative branch of study has focused on the language used in abstracts and compared this language use between disciplines. Liu and Zhu (2023) studied the use of linguistic positivity, that is, the use of certain words categorised as being either positive or negative, in eight academic disciplines that were themselves categorised as either hard (Aerospace, Automation & Control, Software Engineering and Transportation) or soft (Communication, Linguistics, Political Science and Sociology). They found that the disciplines categorised as hard showed a higher and faster-growing use of positive vocabulary in comparison to the softer disciplines. Edlinger et al. (2023), however, compared abstracts from the disciplines of Psychology, Biology and Physics in their use of valence-loaded scientific jargon and found that Psychology featured more of this positive language use than the other two harder disciplines. Edlinger et al. suggest that these seemingly contradictory findings may be due to the differing methodologies of the two studies. Indeed, it may also be due to the different disciplines under study.

Darabad (2016) studied abstracts from Applied Linguistics, Applied Mathematics and Applied Chemistry. He predominantly focused on the rhetorical structure but also considered the authors' self-mention, voice, and tense of the verbs used. The study found the highest use of passive in Applied Chemistry (48% of abstracts) compared to Applied Mathematics (29%) and Applied Linguistics (23%). All three disciplines tended to prefer present tense over past tense, with Applied Mathematics showing the strongest tendency (83% of the abstracts used present tense) compared to 75% for Applied Chemistry and 67% for Applied Linguistics. Darabad also found that Applied Linguistics and Applied Mathematics commonly used personal and possessive pronouns such as *I*, *we*, *our*, etc. whereas pronouns were infrequent in Applied Chemistry abstracts.

Ebrahimi and Chan (2015) used RA abstracts in the disciplines of Applied Linguistics and Economics and compared the discourse functions of grammatical subjects in four domains, the most relevant to the current study being the discourse participant subcategory of the participant domain, which was indicated by the use of *I* and *we*. Their results showed a difference between the two disciplines in the application of the discourse participant category. Applied Linguistics papers used it in only 7% of cases, while Economics papers used it in 29% of the grammatical subjects identified. The authors concluded that "a higher reliance on this GS in Economics RA abstracts could make the texts more interactional and argumentative" (p 387). In a follow-up study from 2022 co-authored with Farnia, Ebrahimi again investigated the grammatical subject, this time in the four disciplines of Applied Linguistics, Economics, Agriculture and Applied Physics. If we again focus on the discourse participant domain, it was most commonly used in Economics, in which 32% of the

grammatical subjects were from this domain compared to only 11% in Applied Linguistics and 6% in Applied Physics.

To summarise the findings, Hyland (2001) as well as Ebrahimi and Farnia (2022) found soft disciplines more likely to use first-person pronouns, but Fanelli and Glänzel (2013) and Darabad (2016) found what could be described as a “U-shaped” relationship, with disciplines at the softer and harder ends of the HoS closer to each other than to those in the middle. Furthermore, Darabad (2016) reported that present tense and passive voice were used more in harder sciences (the latter, however, also in a somewhat inverted “U-shaped” relationship). These results echo the “U-shaped” relationships that were also found in the study by Benjafield (2020). Intriguingly, a large-scale quantitative study by Dong et al. (2024) also reports inverted U-shaped patterns in the variation of linguistic features across disciplines ordered by the HoS. Central disciplines such as Biology and Chemistry exhibited peak values for several linguistic indicators, including punctuation frequency, unique word count, total word count, sentence count, varied sentence length dispersion and the Dale-Chall and SMOG readability indices, with disciplines at the extremes showing reduced values. This finding further challenges a simple linear interpretation of the hierarchy of sciences. These findings clearly suggest that the relationship between disciplines in the HoS and their linguistic features may not form a clear linear relationship.

Although some of the articles discussed have looked at a broad range of features related to the text, all have focused on, at most, a very limited selection of linguistic features supposed to reflect the hardness or softness of disciplines and/or a limited choice of disciplines that did not always correspond to that of the HoS. It would therefore be informative to first identify the features characteristic of the language used in published research articles overall, and then to see whether this bears any relationship to the order of the disciplines on the HoS.

One method particularly suited to surveying a broad range of linguistic features and identifying the most salient ones is Biber’s Multidimensional approach (Biber, 1988). This is a corpus-based method of linguistic analysis that identifies patterns of co-occurring grammatical and lexical features to reveal underlying dimensions of variation across different types of texts and speech. This method has been used to investigate linguistic variation in academic research articles in numerous publications. The most notable of these is Gray (2015), which uses corpus-based multidimensional analysis to show that variation in academic research writing is shaped not only by discipline but also by research purpose, methodology, and communicative function.

Taking a more specific focus, Jin’s (2018, 2021) studies used Biber’s multidimensional approach to identify recurring linguistic patterns in discussion sections of chemical engineering research articles and compare high- vs low-impact papers. She found six key dimensions, including: Involvement and interactivity; Narration vs non-narration; Further explanation and elaboration on evaluative expressions; Informational density; Stating results/claims; and Expression of denial relationships in experimental findings, with significant differences in how these are used across impact levels. The 2021 study reinforced these findings with expert interviews, highlighting how stance and evaluation strategies shape effective academic writing. Lan et al. (2026) showed that MDA can effectively capture subtle differences between closely related disciplines, revealing that even adjacent fields like economics and marketing have clearly distinct multidimensional linguistic profiles.

This paper builds on such work in its attempt to identify and describe the linguistic characteristics of RAs by using Biber’s multidimensional approach. Thus, instead of choosing linguistic features thought to be harder or softer and measuring them, this study

will begin by identifying the typical characteristics of all RAs and then use those features to investigate whether there is a systematic trend across the hard/soft continuum. In a second step, we will identify which disciplines are linguistically most similar to each other, order them accordingly and then compare this order to the traditional HoS. Finally, we investigate whether disciplines considered ‘harder’ in the traditional hierarchy indeed tend to employ grammatical features more commonly linked to characteristics of ‘hard’ sciences and, inversely, whether disciplines traditionally considered ‘softer’ tend to employ features commonly attributed to them. Based on the previous literature, these softer grammatical features would include more use of first-person pronouns and less use of passive voice, while harder sciences would express themselves more positively and use more present tense.

Methods

Corpus

Our corpus of 39,809 abstracts comprised 7,630,993 words from RAs published in 2021 and 2022. The reason for focusing on the language used in abstracts are an increasingly vital part of an RA as they summarise and showcase the content of the article. Although more data would have been available in full texts, many RAs were not freely available for copyright reasons in the country in which the research was conducted. Nonetheless, abstracts still display a representative sample of the language used in RAs as they usually provide information about each of the different sections of the article. The abstracts and their associated bibliographic information were downloaded from the Web of Science Core Collection online database of reference and citation data for academic journals (Clarivate, 2023).³

The only search criteria used were publication year, language (English) and document type (article). This resulted in over 2.3 million records for 2021 and 2.4 million for 2022. With the results sorted by relevance, the full records of the first 20,000 RAs were exported for both years. From this initial 40,000 RAs, 191 were excluded as they contained no abstract data. Table 1 shows the number of RAs from each broad category. Note that each RA may be categorised into up to four broad categories to allow for multidisciplinary research. Although the percentage of Arts and Humanities articles (0.3%) was very small, the absolute number of 156 RAs was in fact representative of the percentage of articles from this category published in the period under investigation.

Multidimensional analysis

A multidimensional analysis (MDA) of linguistic variables was performed following the method suggested by Biber (1992). The MDA was performed on the full corpus of 39,809 abstracts in order to get a full picture of the grammatical co-occurrences in the language

³ The records were downloaded from WoS prior to Clarivate’s 2024 reorganisation of categories. Further information on the WoS categories can be found at: <https://webofscience.zendesk.com/hc/en-us/articles/27505726032017-Web-of-Science-Subject-Categories> and <https://webofscience.zendesk.com/hc/en-us/articles/38543541713169-Research-Areas>

Table 1 Multidimensional analysis loadings

Dimension 1		Dimension 2	
Authors present		Passive voice	
Verb present tense	0.814	Verb past participle	0.910
Determiner	0.459		
First person pronouns*	0.320	Agentless passive*	0.716
(Symbol)	−0.317	(First Person pronouns*)	−0.310
Independent clause coordination*	−0.321		
Type token ratio	−0.393		
Other noun*	−0.535		
Cardinal numbers	−0.600		
Verb past tense	−0.668		
Dimension 3		Dimension 4	
Hedging		Non-mathematical description	
Verb, base form*	0.904	Average word length	0.786
To-infinitive	0.676	Attributive adjectives	0.330
Possibility modal*	0.327	Nominalisation*	0.320
		Symbol	−0.342
		BE as main verb	−0.354
		(Cardinal numbers) ^a	−0.470

*These features are explained in Appendix 1

^aThe MAT Tagger also tags decimal numbers as cardinal numbers

used in RA abstracts. Before the MDA, any incidental text, such as copyright information, was removed, and then the texts were grammatically tagged using the open-source software Multidimensional Analysis Tagger (MAT) (v. 1.2 Nini, 2019). The MAT tagger tags each token with at least one grammatical tag from a list of 42 tags. It then outputs a frequency score for each tag on each text in addition to the number of tokens, the average word length and the type/token ratio for each text. Next, we conducted an exploratory factor analysis (EFA) to extract factors from the MAT tagger data. The EFA used the *psych* package (Revelle, 2026) and was carried out in R (v4.5.0 R Core Team, 2025). A visual inspection of the scree plot, the interpretability of the resulting factors, the deflection points of the eigenvalues and the Tucker-Lewis index of factoring reliability were all considered in order to determine the best number of factors. The “Varimax” factor rotation was used to force each dimension to load on as few factors as possible. Only features with a loading higher than 0.3 were included in the dimension. Furthermore, if a feature loaded on more than one dimension, the feature was included in both dimensions, with the lower loading shown in brackets. The obtained dimensions list the co-occurring linguistic features with a weighting ranging from -1 to 1. The dimensions are then given a descriptive label based on these linguistic features and their weightings. The mean dimension score for each RA was calculated by multiplying the z-score of the frequency of the linguistic features on a dimension by the weighting and calculating the mean of these weighted scores.

The analysis of the MDA involved deciding how many dimensions to select, interpreting the features in those dimensions and using this interpretation to give them a descriptive label. To aid in this process, the most common collocates of salient features were extracted

from the data using AntConc (Anthony, 2024). We decided upon four dimensions, as shown in Table 1.

Dimension 1

The first dimension, which we have labelled Authors Present, is characterised by the frequent use of a present tense verb with a first-person pronoun and a determiner, as well as by a marked tendency to avoid past tense verb forms and cardinal numbers. The most common token before a present-tense verb is *we*. The most common examples of this are *we find*, *we propose*, *we show*, *we have*, *we present* and *we consider*. The most common grammatical tag after a present tense verb is a determiner, the most common of which is *the* followed by *a*. The most common grammatical tag after any first-person pronoun (*I*, *we*, *me*, *us*) is present tense verb. The most common present tense verb in this case is *present* followed by *use*, *examine* and *study*. Representative examples of this pattern are:

- (1) I then present Critical Realism’s process of critique and explanation, how they connect to human emancipation and how they can lead to impactful change in autism research... (Psychology).
- (2) Given its popularity, we examine how this methodology approaches the interpretive process of fieldwork. (Psychology).
- (3) In particular, we find the mod 2 cohomological invariants for the symmetric groups... (Mathematics).
- (4) Hence, our study suggested that M.s-5Y may have potent efficacy in ameliorating NAFLD. (Chemistry).
- (5) In this study, we use the integrated concept of the historical water supply facility to advance the regional characterization of point-like facilities. (Arts & Humanities).

We have chosen the three-way ambiguous dimension label “Authors Present” as this captures the three key features of Dimension 1: Firstly, the authors are ‘present’ (adjective), i.e. the use of first-person (typically plural) pronouns explicitly refers to the author(s). Secondly, the label refers to ‘present tense’, in which the verbs figure almost without exception. Thirdly, the most common (present tense) verb is in actual fact *present*. A prototypical example combining all three characteristics is therefore (6) below.

In this work, we present a methodology for the prediction of SEPs from the soft X-rays of solar flares... (Physical Sciences; Astronomy & Astrophysics).

Dimension 2

The second dimension, which we have labelled Passive Voice, features, on the one hand, the common use of past participles and agentless passive constructions, and a dispreference for first-person pronouns, on the other. This seems plausible as (in particular, agentless) passives are often used in academic writing to avoid first-person pronouns (like *I*, *we*, *our* etc.) and this strategy thus creates some rhetorical distance between the author(s) and their research, often in an endeavour to express objectivity.

The most common words tagged as agentless passives were the verbs *were*, *was* and *is*. The most common grammatical tag immediately after an agentless passive was a past participle verb. The most common past participles following an agentless passive were *used*,

found and *associated*. Likewise, the most common grammatical tag before a past participle was an agentless passive. Some representative examples of this pattern are:

- (7) It was used to construct this sensor for quantification of these important pesticides and sensitive determination of metribuzin. (Chemistry).
- (8) Interestingly, chronotype was found to be positively associated with emotional functioning... (Biology).
- (9) In Study 1, coping self-efficacy, but not general self-efficacy, was associated with information avoidance. (Psychology).
- (10) Foliage spray and root irrigation experiments were conducted under laboratory conditions. (Chemistry).
- (11) Shear loading tests and FE analysis were conducted to investigate this phenomenon. (Arts & Humanities).

Dimension 3

We have labelled the third dimension Hedging and it includes the features typical of this rhetorical strategy, most obviously the use of modal verbs expressing possibility (*can*, *may*, *might*, *could*; see Biber, 1988), which combine with base forms of the verb. Possibility modals were most commonly followed by forms such as *help*, *provide* and *have*. Very prominent are also complex verb phrases that include *to*-infinitives, such as *help to explain/understand/ identify*. These *help*-constructions generally have an additional downtoning effect. Typical examples of dimension 3-type hedging by the authors are the following:

- (12) In addition, Gronberg et al. tested thousands of compounds to see if they could block the activity of the protein, identifying two that could kill bacteria. (Biology)
- (13) This work demonstrated that nAl@F-x nanocomposite prepared by PFOA coating on nAl could effectively solve the detrimental effect of in situ formed Al₂O₃... (Chemistry)
- (14) but also might help to develop the fast synthesis methods for other aluminosilicate zeolites. (Chemistry)
- (15) by examining how the end of official school segregation may have helped to shape the racial contours of nineteenth-century urban development. (Arts & Humanities)
- (16) A greater understanding of such patterns could help inform the way healthcare professionals engage in conversations about end-of-life care planning with older LGB people. (Sociology)
- (17) The current study would help to understand the impact of gender on work-family interface... (Psychology)
- (18) Although it may be a matter of perception for Q-followers, a sense of precarity is enhanced by QAnon... (Sociology)

Dimension 4

Finally, Dimension 4, Non-Mathematical Description, shows the use of attributive adjectives and nominalisation, which naturally results in longer words, and thus, average word length is another feature of this dimension. Some typical examples are:

- (19) there have been many studies on renal IRI, although an effective treatment method has not been developed. (Biology)
- (20) In the research about teachers' professional development, there is an increasing interest in analyzing the processes... (Psychology)
- (21) Sufficient conditions for Deza graphs to be divisible design graphs are given... (Mathematics)

As the label implies, it is the dispreferred features that are the most salient in this dimension. Disciplines which score highly on this dimension are those that display a relative lack of symbols and cardinal numbers. Figure 1 shows the average word length (mean length of the words in the text in orthographic letters) for each discipline, with Astronomy & Astrophysics displaying the lowest average word length and Psychology the highest. Of the dispreferred features, typically the combination of *be* as the main verb with a cardinal (or decimal) number and/or a symbol was used to present data as in the example:

- (22) The pulse duration behind the anode is about 100 ps. (Physics)
- (23) The hit rate (DS) of the FEPA model decomposition method is 72.22%, 10.86% higher than the EMD-BPNN model. (Mathematics)
- (24) The Dice coefficient was 0.918 at the pixel level, and there were no false positives at the slice level using U-Net++ with 9 input slices. (Biology)

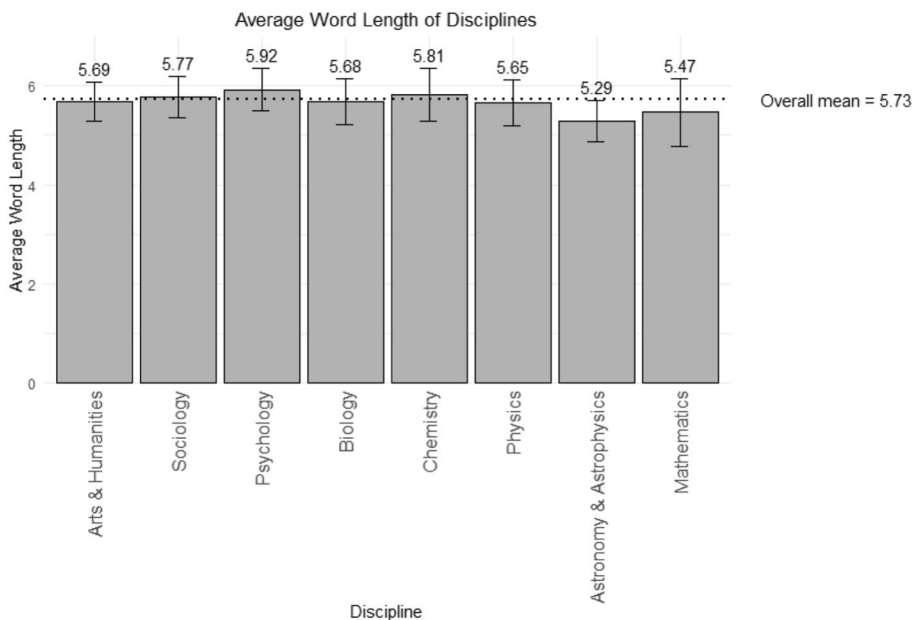


Fig. 1 Average word length of each discipline. The centre line represents the mean average word length of all RAs from the full corpus, not just those from the HoS disciplines

Discipline choice and categorisation

The subsequent analysis of the language of the HoS disciplines was then conducted on subsets of the larger corpus as described above. These sub-corpora were chosen using bibliographic data attached to each WoS record. The original HoS disciplines included in Comte's text from 1853 include Sociology, Biology, Chemistry, Physics, Astronomy and Mathematics. Taking our lead from previous research into this topic (e.g. Simonton, 2015; Smith et al., 2000), we have expanded this selection slightly to include Arts & Humanities and Psychology, resulting in eight disciplines that will be investigated in detail. The complete records for each RA download from WoS contained a variety of bibliographic information, including the WoS broad category, research area and subject category for that article. These are bibliographic tags identifying the subject matter of each article. There are five broad categories: Arts & Humanities; Social Science; Life Sciences & Biomedicine; Physical Sciences; Technology. Research area is a more general descriptor, and the subject category is the most specific, with each subject category being mapped onto one research area. Each RA is categorised with between one and five of the 139 possible research areas and between one and three of the 254 possible subject categories. For the purposes of this study, the first listed research area was used to select abstracts for each discipline sub-corpus. It is assumed that the first research area listed is the most representative of the content of the article. In order to decide which RAs belonged to disciplines in the HoS, we selected the bibliographic tags that most closely matched those used in previous studies concerning the HoS. In the case of Chemistry and Physics, these tags were assigned in the WoS at the level of *research area*. Astronomy & Astrophysics, Mathematics, Psychology and Sociology were all assigned at both a *research area* and WoS *category* level. In these cases, all abstracts assigned the relevant tags at the *research area* level were included in this study. Only for the discipline of Biology was there no research area with that name. In this case, abstracts tagged Biology at the WoS *category* level were used. In the case of Arts & Humanities we have used all abstracts categorised as such under this broad category. Table 2 shows the number of abstracts that were tagged as belonging to HoS disciplines.

Table 2 Total number of abstracts from each HoS discipline

Broad category	Number of abstracts
Arts & humanities	226
Sociology	108
Psychology	1805
Biology	237
Chemistry	3779
Physics	741
Astronomy & astrophysics	219
Mathematics	1093

Hypotheses and research questions

In order to determine if there is grammatical evidence for the HoS and if disciplines that are closer to each other on the HoS are also more similar in their use of grammatical features, firstly the Euclidean distance between the data points was calculated using the `dist()` function in R. Next, we used non-metric multidimensional scaling (NMDS) stress test to decide how many dimensions would best represent the data. Finally, the data was analysed using multidimensional scaling with the `cmdscale` function in R to plot the given number of dimensions (v4.5.0; R Core Team, 2025). The technique of hierarchical clustering, also known as hierarchical cluster analysis, was used to identify which disciplinary categories are most similar to each other using the Mean Dimension Scores for each of the HoS disciplines. This was performed in R with the `hclust` function using Ward’s minimum variance method.

Finally, we can test if the use of grammatical features supports the idea that some sciences are harder (i.e., more impersonal, certain and technical) than others. Having identified four prominent dimensions of grammatical co-occurrence patterns in the corpus of RAs, it is possible to make some predictions as to how prevalent the use of these patterns are in different disciplines across the HoS. Regarding Dimension 1, Authors Present, Fanelli and Glänzel (2013) state, when discussing their expectation that harder sciences should use fewer first-person pronouns, “scientists aim at making universal claims, and their style of writing tends to be as impersonal as possible.” This aligns with the assumption of the HoS that harder sciences are more objective. Thus, hypothesis one is that harder sciences should score lower on Dimension 1, Authors Present, and higher on Dimension 2, Passive Voice. According to Fanelli, another central assumption of the HoS is that there is more consensus in the harder sciences, “on the relevance of findings, on the correct methodologies to use, even on the relevant research questions to ask, and therefore ultimately on the validity of any particular theory” (2019). Li et al. (2025) also allude to this, but from the perspective of readers from the harder disciplines. Whether or not this is the case is the subject of debate but if so, this proposed lack of consensus in the softer sciences due to the increased complexity of its subject matter could be reflected in higher scores on Dimension 3, Hedging, as this is a typical way in which an author can express a lack of certainty. This could also be a reflection of the proposed increased linguistic positivity in the harder sciences (Edlinger et al., 2023; Liu & Zhu, 2023). Finally, Dimension 4, Non-mathematical Description, which has as its most obvious features the lack of numbers and symbols, would routinely be associated with softer sciences. The four dimensions and their related hypotheses are summarised as follows:

Dimension 1 Authors Present should be more frequent in the softer sciences.

Dimension 2 Passive Voice should be more frequent in the harder sciences.

Dimension 3 Hedging should be more frequent in the softer sciences.

Dimension 4 Non-mathematical Description should be more frequent in the softer sciences.

In order to test these hypotheses, Spearman’s rank correlation was calculated as this is the most appropriate measure of association when one variable is continuous, the other is ordinal, and the ordinal variable has more than five or six levels (Khamis, 2008).

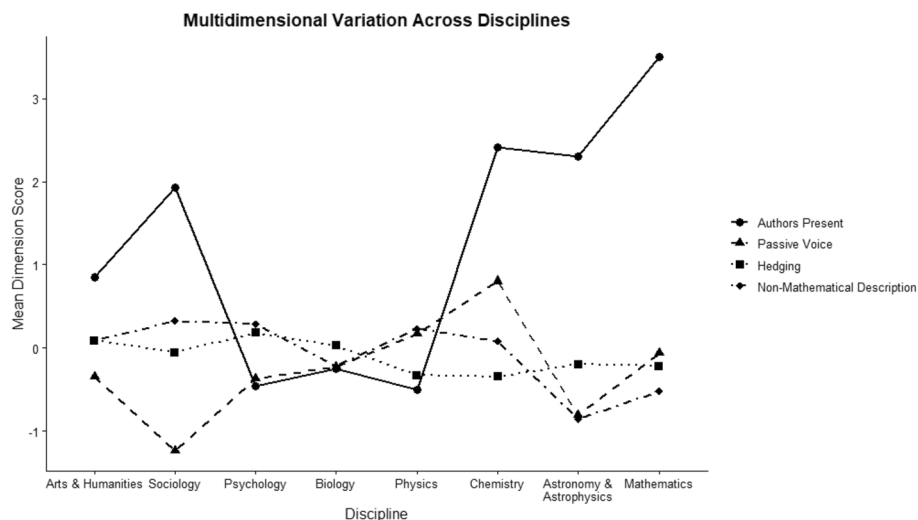


Fig. 2 Mean Dimension Scores of each of the Disciplines of the Hierarchy of Sciences in order of softest to hardest (The order of hardness is based primarily on Comte's (1853) classification with the addition of Psychology in the position hypothesised by Simonton (2015) and Arts & Humanities in the softest position.)

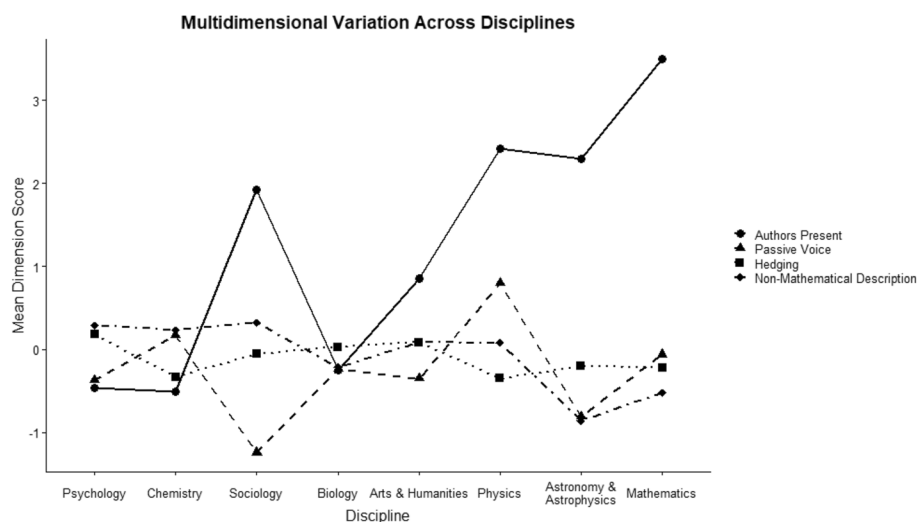


Fig. 3 Mean dimension scores of the disciplines of the hierarchy of sciences ordered by Euclidean distance

Results

After identifying the four linguistic dimensions described in the previous section, we were able to calculate mean dimension scores for RAs which shared a primary subject categorisation in WoS. Thus, for every RA categorised as, for example, Physics, we calculated the mean score on each dimension for those RAs. The Mean Dimension Scores for each

discipline of the HoS on each dimension are shown in Fig. 2. In this case, the data is presented in order of perceived softness to hardness. In Fig. 3, the same data is presented, but the disciplines are ordered according to their similarity calculated by Euclidean distance. Ordering the disciplines in this way shows a notable difference to the traditional order, although the harder end of the spectrum is largely intact.

A Spearman rank correlation was used to calculate whether there was a relationship between the mean dimension scores for each of the disciplines and their proposed order according to the HoS. The results show that each of the dimensions was strongly correlated with the HoS: Dimension 1 (Authors Present) $\rho=0.548$, $p=0.171$; Dimension 2 (Passive Voice) $\rho=0.619$, $p=0.115$; Dimension 3 (Hedging) $\rho=-0.596$, $p=0.132$; Dimension 4 (Non-mathematical Description) $\rho=-0.666$, $p=0.083$.

Figure 4 shows a dendrogram of the results of the hierarchical cluster analysis. Visualising the data in this way broadly supports the HoS in suggesting at the top level a division between hard and soft sciences, with Mathematics, Astronomy & Astrophysics and Physics on the hard side with the unexpected inclusion of Sociology, and Arts & Humanities, Psychology and Biology clustered on the soft side with the unexpected inclusion of Chemistry. The finest level of clustering pairs Psychology with Biology.

Figure 5 shows the plot of the NMDS stress test. A score <0.05 is considered “excellent”. Between $0.1 - 0.2$ is considered “fair” and 0.000 typically indicates overfitting issues. The actual scores are: one dimension: 0.138 ; two dimensions: 0.024 ; three dimensions: 0.000 ; four dimensions: 0.000 . From this, we concluded that two dimensions indicate the best way to express the data with a good fit and without any overfitting.

Figure 6 presents the distances between the mean dimension scores of the disciplines in a *two-dimensional* space, allowing us to represent a more nuanced model of the relative similarities between disciplines. Although, *prima facie*, there does not seem to be a clear separation between the perceived hard and soft subject areas, Fig. 6 shows that the soft sciences (represented as circles) are all in the bottom-right quadrant, while the harder sciences (triangles) are spread across the remainder of the space.

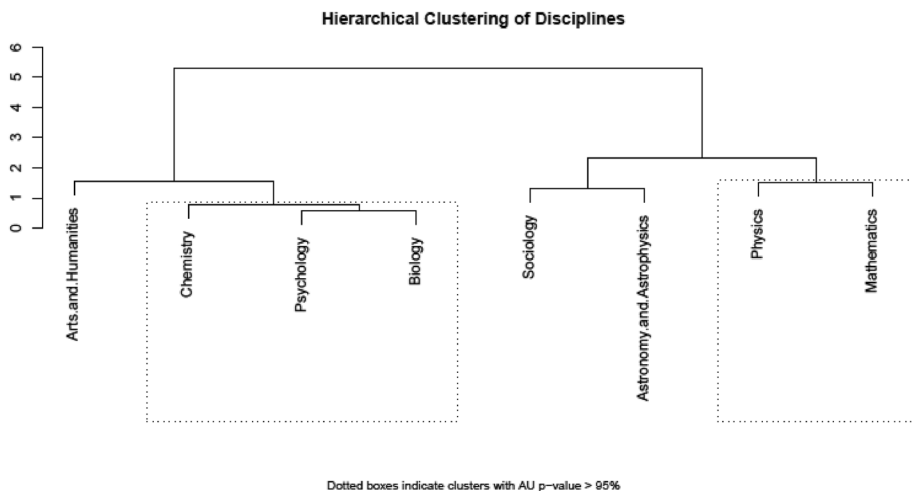


Fig. 4 Cluster Dendrogram of the HoS subjects according to their mean dimension scores

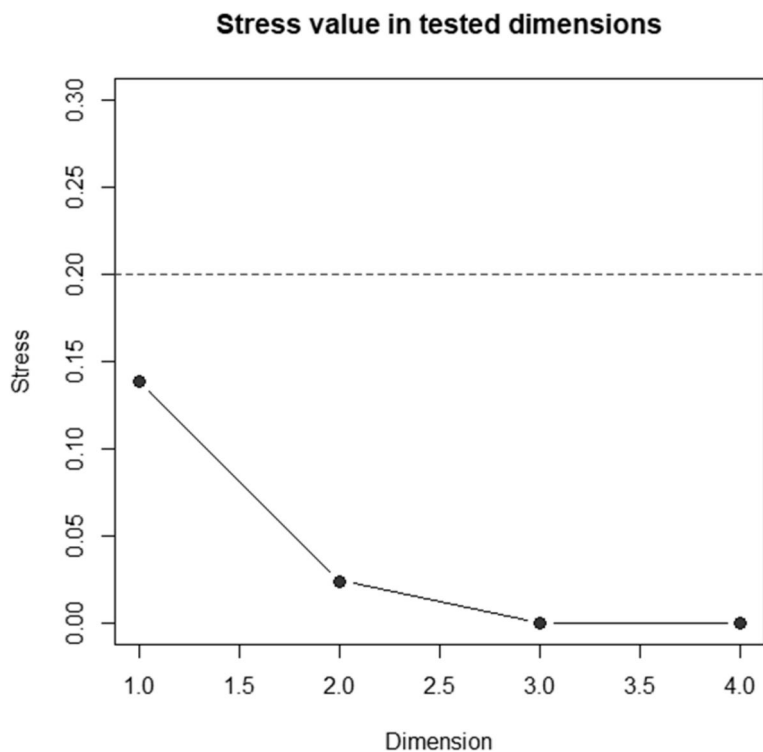


Fig. 5 NMDS stress test

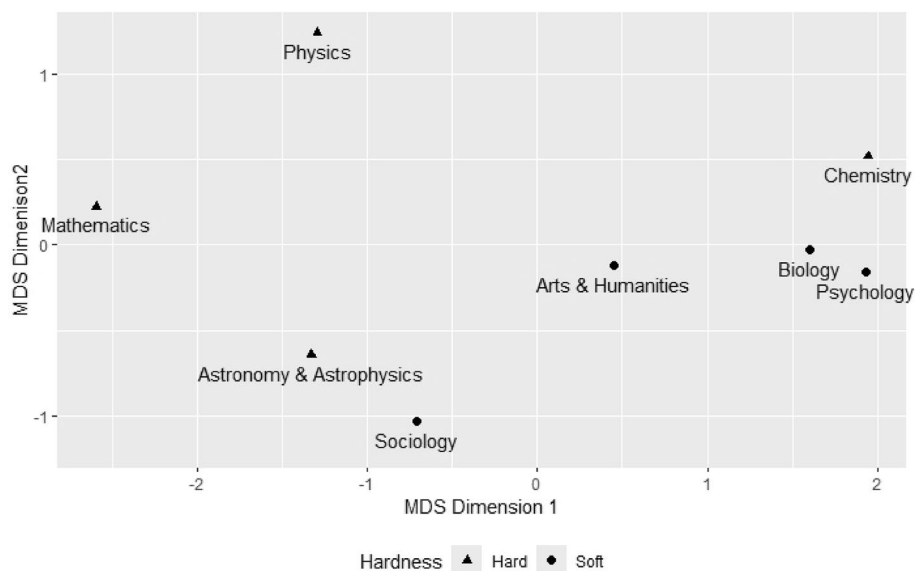


Fig. 6 Distances between the HoS disciplines in a two-dimensional space – hardness assigned after Comte's designation as organic (soft) or inorganic (hard)

Discussion

The present study investigated whether the widely held characteristics of academic disciplines in the HoS are reflected in the language used by authors in published research articles. It identified four prominent dimensions of grammatical patterns, namely Authors Present, Passive Voice, Hedging and Non-mathematical Description. Across previous multidimensional studies, our four dimensions align broadly with established patterns but also show important refinements. Gray's (2015) dimensions similarly capture involvement, procedural discourse, agency, and abstract "academese," but our Authors Present dimension represents a more narrowly operationalised form of involvement focused on explicit authorial self-mention and present-tense rhetorical acts, while our Passive Voice dimension resembles Gray's procedural description pole. Jin's (2018, 2021) studies of RA discussion sections similarly identified Involvement and interactivity; Narration vs non-narration; Further explanation and elaboration on evaluative expressions; Stating results/claims; Informational density; and Expression of denial relationships in experimental findings, with clear overlap with our dimensions of Narration vs non-narration (Authors Present), Informational density (Non-mathematical description), though Jin additionally foregrounds other functions not separated in our model.

Hypothesis 1, that Authors Present should be more frequent in the softer sciences, was not supported as Physics, Astronomy & Astrophysics and Mathematics had the highest scores on this dimension. However, Arts & Humanities and Sociology, the softest sciences, clearly used this dimension more than the more 'central' sciences of Psychology, Biology and Chemistry. Hypothesis 2, that Passive Voice should be more frequent in the harder sciences, was generally confirmed, with the notable exception of Astronomy & Astrophysics, which showed the second lowest score on this dimension. Hypothesis 3, that Hedging should be more frequent in the softer sciences, was also supported, although the differences were not large. Finally, regarding Hypothesis 4, that Non-mathematical Description should be more frequent in the softer sciences, it was certainly the case that Astronomy & Astrophysics and Mathematics showed a greater use of mathematical description, but amongst the other disciplines there was very little difference.

Although there is some support for the trends expected according to the literature, there are also several unexpected findings. It is particularly noteworthy that the first two dimensions (which are the most significant as they capture the largest share of the linguistic variation) are the ones that seem to show a "U" shape when the disciplines are ordered according to the HoS. This pattern is consistent with recent large-scale evidence reported by Dong et al. (2024), who also identified inverted U-shaped distributions of multiple linguistic features across disciplines. This suggests that the disciplines at the hardest and softest extremes (Mathematics, Astronomy & Astrophysics, Arts & Humanities and Sociology) have more in common than was conventionally assumed. The cluster analysis, which is visualised in the dendrogram in Fig. 4 also shows a general support for the hard/soft distinction in academic disciplines. But again, there are discrepancies. Most notably, Chemistry is clustered with the soft sciences. However, it is most closely clustered (AU p-value > 95%) with Biology and Psychology. This hints at a further grouping of 'firm' sciences between soft and hard (by analogy with *firmware's* intermediate position between *software* and *hardware*). It is also worth noting that these results generally support the findings of Simonton (2015), who attempted to place Psychology in the HoS using a variety of measures as described in the introduction to this paper. On each of the dimensions, Psychology scores close to Biology on the mean dimension scores.

Given the discrepancies we have discussed, our study highlights one weakness of the venerable HoS – its age. It is highly likely that cultural shifts would occur within discourse communities, resulting in the altered usage of highly salient and often contentious grammatical features such as passive voice. And these changes are possible in much shorter timeframes than the roughly 200 years since the HoS was first proposed. In fact, just such a change was found by Jiang and Hyland (2022), who found a decrease in the use of the past tense and usage of passive in abstracts in the harder sciences (Biology and Engineering) between 1990 and 2020.

Another interesting and surprising finding regards the use of Dimension 1, Authors Present, especially the use of first-person pronouns. Fanelli and Glänzel (2013) found the use of this feature to be one of the few that did not support the HoS amongst the bibliographic measures they looked at. Their data, especially for first-person singular pronouns, was largely in accordance with ours. In both cases, Mathematics and Humanities had more in common with each other than with Biology. However, if one considers that passive voice correlates negatively⁴ in linguistic practice with first-person pronouns (as authors often use passive voice to avoid subjects *I* or *we* in active constructions), then Fanelli and Glänzel (2013) might have found support for their hypothesis on this item also, had they used passive voice instead of first-person pronouns as their variable. Our data shows a clear trend for the increased use of passive voice towards the harder sciences, at least when focusing on Sociology, Psychology, Biology, Chemistry and Physics.

Both Fanelli and Glänzel's (2013) and our findings contradict those of Hyland (2001), who also looked at self-mention in RAs, comparing disciplines grouped into hard and soft categories, and found that the soft sciences more frequently use self-mention. The crucial and explanatory difference seems to be in the selection of disciplines. Hyland did not include Mathematics and Space Science (as do Fanelli & Glänzel, 2013) or Astronomy & Astrophysics (as does the current paper), which yielded unexpected results in both papers.

These contradictory findings open a proverbial can of worms regarding the issue of authors referring to themselves in RA texts as advice from style guides, requirements from journals and expectations based on a belief in the ideas behind the HoS are often in conflict. Fanelli and Glänzel (2013) based their hypotheses on the suggestion that, “scientists aim at making universal claims, and their style of writing tends to be as impersonal as possible.” Even Albert Einstein, probably, in the popular imagination at least, the most famous, iconic and popular scientist ever, wrote that when “talking about scientific subjects, the little word ‘I’ should play no part in ... expositions” (Einstein, 1934, p.113). Conversely, the journal *Nature*, one of the most famous scientific journals, advises potential authors, “to write in the active voice (‘we performed the experiment...’) as experience has shown that readers find concepts and results to be conveyed more clearly if written directly” (Springer *Nature*, n.d.). Perhaps fittingly for a discipline in the middle ground of the HoS spectrum, the American Society for Microbiology takes a mixed approach in its recommendation in its style manual, “passive voice sometimes functions well in the methods section. Elsewhere in a scientific paper, however, it rarely should be chosen. Why, then, do scientists use so much passive voice? Perhaps this bad habit results from the erroneous idea that it is somehow impolite to use first-person pronouns” (ASM Journals, 2020). Perhaps then, the unexpected results for the Authors Present dimension reflect the confusion over how to appropriately use this relatively salient grammatical feature.

⁴ A correlation analysis was run on the complete corpus of 39,809 abstracts collected. It found a weak negative correlation between the use of first-person singular and passive voice ($r = -0.305$, $p < .001$).

A further criticism we would direct towards the HoS is that it is one-dimensional. Its simplicity, while probably being partly responsible for its enduring popularity, reduces any possible understanding of the relationships between disciplines to a linear one. This limitation was already addressed in Fig. 6, in which the linguistic relationships between disciplines are rendered through two dimensions. When viewed this way, the similarities between the HoS disciplines begin to resemble not so much a linear hierarchy as an academic space. While some of the generally assumed distinctions are present (e.g. Sociology is distant from Physics) and similarities (Psychology is adjacent to Biology), we now see some unexpected similarities such as that of Sociology and Astronomy & Astrophysics. This new representation recalls Benjafield's (2020) finding of a high degree of vocabulary sharing between Mathematics, Philosophy and Linguistics, suggesting the areas at the extremes of the HoS are not as distant from each other as is commonly assumed.

Of course, with the addition of more disciplines into any future analysis, a more detailed pattern is likely to emerge; nonetheless we believe that a *space* of the sciences would be a more accurate and informative conceptualisation. Furthermore, this space of the HoS disciplines is less condescending towards those colleagues deemed to be at the “bottom” of the hierarchy and perhaps less controversial overall. The higher complexity of Fig. 6, which provides, in a two-dimensional MDS-generated space, important similarity estimates through the measurable distances between disciplines, offers more valuable information and conforms better to actual writing practices than a simplistic, once-positated, and rightfully controversial, one-dimensional hierarchy. Needless to say, further research is needed that gives a more nuanced picture of the relationships between academic disciplines.

One criticism that could be levelled against the current paper is that the corpus was compiled only from abstracts. While a corpus compiled from full texts may be somewhat richer and more varied, abstracts typically summarise each section of an RA and thus contain language representative of the whole article. In fact, the use of abstracts has advantages that we believe outweigh its limitations. Firstly, the smaller text size allows a wider selection of authors and journals to be included in a corpus of the same size, thereby giving, in fact, a more representative sample, and one that consists of sufficient number of RAs even for the least frequent discipline. Secondly, abstracts are an increasingly important part of a RA as they perform both a summarizing and a persuasive role.

Another potential criticism is that the corpus was restricted to abstracts published in 2021–2022 and therefore the present study does not address possible diachronic shifts in disciplinary writing practices. Although beyond the scope of the present study, this remains an important direction for future research.

A final limitation resulted from focusing on the HoS as a theoretical construct. The disciplines in the HoS are only a small selection of all academic disciplines. The disciplines which are traditionally included in the HoS, even with the inclusion of Psychology, only comprise 8,696 from the 39,809 abstracts collected (21.8%).⁵ Obviously there is a great deal of scope for further research into how other disciplines fit relative to the HoS scheme.

⁵ The full corpus consisted of RAs from all disciplines published in the selected time period. The multidimensional analysis was performed on this full corpus. We then used the data from each of the HoS disciplines for our comparison.

Conclusion

The HoS is well established in the minds of the general public and many academics for good reason. It is simple and intuitive. Furthermore, a number of studies have found some support for the basic ideas relating to a hierarchy with soft(er) and hard(er) disciplines. However, if the so-called harder sciences are in fact more objective and in consensus—as suggested by Fanelli (2010)—then these attributes are not clearly or consistently expressed in the way they communicate their research in published research articles, for instance, through a preference for passive voice and other constructions over first-person pronouns. Even in the use of numbers and symbols, only Astronomy & Astrophysics and Mathematics differ meaningfully from the other disciplines of the HoS investigated in the present study. Given the explosion of new data visualisation techniques that have become available with improved computing power, we believe it is time we did away with a one-dimensional hierarchy and looked at the nuanced relationships between different sciences with greater attention to detail. The complex relationships among the sciences call for multidimensional models, and our two-dimensional representation established for scientific Englishes (Fig. 6) may constitute a first step in that direction.

Appendix

The following explanations of how the grammatical features were tagged are taken from the MAT Tagger manual (Nini, 2019):

Independent clause coordination

This tag is assigned to the word *and* when it is found in one of the following patterns: (1) preceded by a comma and followed by *it, so, then, you, there* + BE, or a demonstrative 31 pronoun or the subject forms of a personal pronouns; (2) preceded by any punctuation; (3) followed by a WH pronoun or any WH word, an adverbial subordinator or a discourse particle or a conjunct. (pp. 30–31)

First person pronouns (FPP1)

Any item of this list: *I, me, us, my, we, our, myself, ourselves*. (p.18)

Nominalizations (NOMZ)

Any noun ending in *-tion, -ment, -ness, or -ity*, plus the plural forms. Although Biber (1988) does not mention that this variable was checked manually, it is likely that a stop list was used to avoid obviously erroneous tagging (e.g. *city*). However, this was not indicated in the appendix of Biber (1988). (p.19)

Total other nouns (NN)

Any noun that has been tagged by the Stanford Tagger as NN and that has not been identified a nominalisation or a gerund is left as such. Plural nouns (NNS) and proper nouns (NNP and NNPS) tags are changed to NN and included in this count. (p.20)

Agentless passives (PASS)

This tag is assigned when one of the two following patterns is found: (a) any form of BE followed by a participle (VBN or VBD) plus one or two optional intervening adverbs (RB) or negations; (b) any form of BE followed by a nominal form (a noun, NN, NNP or personal pronoun, PRP) and a participle (VBN or VBD). This algorithm was slightly changed from Biber's version in the present tagger. It was felt necessary to implement the possibility of an intervening negation in the pattern (b). This

tag is therefore assigned also in the cases in which a negation precedes the nominal form of pattern (b). (p.20)

Possibility modals (POMD)

The possibility modals listed by Biber (1988): *can, may, might, could*. (p.27)

The following explanation of how the grammatical features were tagged are taken from the Part of Speech Tagging Guidelines for the Penn Treebank Project (San-torini, 1990):

Verb, base form (VB)

This tag subsumes imperative, infinitives and subjunctives.

Examples: Imperative: Do (VB) it.

Infinitive: You should do (VB) it.

We want them to do (VB) it.

We made them do (VB) it.

Subjunctive: We suggested that he do (VB) it.

Author contributions Conceptualization: D.W.; Methodology: Don.W.; Formal analysis and investigation: D.W.; Writing—original draft preparation: D.W.; Visualisation: D.W.; Writing—review and editing: D.W., M.K.; CC.C.; Supervision: M.K., CC.C.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anthony, L. (2024). *AntConc* (Version 4.3.1) (Computer Software). Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software/AntConc>
- Benjafield, J. G. (2020). Vocabulary sharing among subjects belonging to the hierarchy of sciences. *Scientometrics*, 125, 1965–1982. <https://doi.org/10.1007/s11192-020-03671-7>
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. (1992). The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities*, 26(5–6), 331–345. <https://doi.org/10.1007/BF00136979>
- Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, 57, 195–203. <https://doi.org/10.1037/h0034701>
- Clarivate Web of Science. (2023) *Research Area (Categories/Classification)*. Retrieved March 22, 2023, from <http://webofscience.help.clarivate.com/en-us/Content/research-areas.html?Highlight=research%20area>

- Cole, S. (1983). The hierarchy of the sciences? *American Journal of Sociology*, 89(1), 111–139. <https://doi.org/10.1086/227835>
- Comte, A. (1853). Characteristics of the Positive Method in its Application to Social Phenomena. In H. Martineau (Transl.), (1896), *The Positive Philosophy of Auguste Comte* Retrieved from: <https://ia902902.us.archive.org/10/items/positivephilosop01comt/positivephilosop01comt.pdf>
- Darabad, A. M. (2016). Move analysis of research article abstracts: A cross-disciplinary study. *International Journal of Linguistics*, 8(2), 125. <https://doi.org/10.5296/ijl.v8i2.9379>
- Dong, S., Mao, J., Ke, Q., & Pei, L. (2024). Decoding the writing styles of disciplines: A large-scale quantitative analysis. *Information Processing & Management*, 61(4), Article 103718. <https://doi.org/10.1016/j.ipm.2024.103718>
- Ebrahimi, S. F., & Chan, S. H. (2015). Research article abstracts in applied linguistics and economics: Functional analysis of the grammatical subject. *Australian Journal of Linguistics*, 35(4), 381–397. <https://doi.org/10.1080/07268602.2015.1070660>
- Ebrahimi, S. F., & Farnia, M. (2022). “We examine the relationship...”: Grammatical subject in research article abstracts: Disciplinary and functional analysis. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, 10(1), 99–118.
- Edlinger, M., Buchrieser, F., & Wood, G. (2023). Presence and consequences of positive words in scientific abstracts. *Scientometrics*, 128(12), 6633–6657. <https://doi.org/10.1007/s11192-023-04864-6>
- Einstein, A. (1934). *Essays in science*. The Philosophical Library.
- Fanelli, D. (2010). “Positive” results increase down the hierarchy of the sciences. *PLoS ONE*, 5(4), Article e10068. <https://doi.org/10.1371/journal.pone.0010068>
- Fanelli, D., & Glänzel, W. (2013). Bibliometric evidence for a hierarchy of the sciences. *PLoS ONE*, 8(6), Article e66938. <https://doi.org/10.1371/journal.pone.0066938>
- Gray, B. (2015). Linguistic variation in research articles: When discipline tells only part of the story. *John Benjamins Publishing Company*. <https://doi.org/10.1075/sc1.71>
- Haslam, N., & Baes, N. (2023). Scientific eminence and scientific hierarchy: Bibliometric prediction of fellowship in the Australian Academy of Science. *Scientometrics*, 128, 6659–6674. <https://doi.org/10.1007/s11192-023-04870-8>
- Holtz, M. (2011). Lexico-grammatical properties of abstracts and research articles. A corpus-based study of scientific discourse from multiple disciplines (Doctoral dissertation, Technische Universität Darmstadt).
- Hyland, K. (2000). Disciplinary Discourses. *Social Interactions in Academic Writing*. London: Longman.
- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20(3), 207–226. [https://doi.org/10.1016/S0889-4906\(00\)00012-0](https://doi.org/10.1016/S0889-4906(00)00012-0)
- Jiang, F (Kevin), & Hyland, K. (2022). Changes in research abstracts: Past tense, third person, passive, and negatives. *Written Communication*, 40(1), 210–237. <https://doi.org/10.1177/07410883221128876>
- Jin, B. (2018). A multidimensional analysis of research article discussion sections in the field of chemical engineering. *IEEE Transactions on Professional Communication*, 61(3), 242–256. <https://doi.org/10.1109/TPC.2018.2817002>
- Jin, B. (2021). A multi-dimensional analysis of research article discussion sections in an engineering discipline: Corpus explorations and scientists’ perceptions. *SAGE Open*, 11(4), Article 21582440211050401. <https://doi.org/10.1177/21582440211050401>
- ASM Journals. (2020) *Writing your paper*. <https://journals.asm.org/editorial-style>
- Khamis, H. (2020). Measures of association: How to choose? *Journal of Diagnostic Medical Sonography*, 24(3), 155–162. <https://doi.org/10.1177/8756479308317006>
- Lan, T., Hu, C., & Lu, C. (2026). A comparative multi-dimensional analysis of economics and marketing research articles. *Journal of English for Academic Purposes*, 81, Article 101690. <https://doi.org/10.1016/j.jeap.2026.101690>
- Li, Z., Shi, Y., & Lei, L. (2025). Metaphor as a springboard to scientific communication: A large-scale study of the use of lexical metaphors across disciplines. *Humanities & Social Sciences Communications*, 12(1), 1–12. <https://doi.org/10.1057/s41599-025-05677-z>
- Liu, X., & Zhu, H. (2023). Linguistic positivity in soft and hard disciplines: Temporal dynamics, disciplinary variation, and the relationship with research impact. *Scientometrics*, 128(5), 3107–3127. <https://doi.org/10.1007/s11192-023-04679-5>
- Munroe, R. Purity. *xkcd*. <http://xkcd.com/435/> (Accessed May 1, 2024).
- Springer Nature (n.d.) *How to write your paper*. <https://www.nature.com/nature-portfolio/for-authors/write>
- Nini, A. (2019). The multi-dimensional analysis tagger. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 67–94). Bloomsbury Academic.
- R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.0) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>

- Revelle, W. (2026). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.6.3. <https://doi.org/10.32614/CRAN.package.psych>
- Santorini, B. (1990). *Part-of-speech tagging guidelines for the Penn Treebank Project* (MS-CIS-90-47). Department of Computer and Information Science, University of Pennsylvania.
- Simonton, D. K. (2015). Psychology as a science within Comte's hypothesized hierarchy: Empirical investigations and conceptual implications. *Review of General Psychology*, 19(3), 334–344. <https://doi.org/10.1037/gpr0000039>
- Smith, L. D., Best, L. A., Stubbs, D. A., Johnston, J., & Archibald, A. B. (2000). Scientific graphs and the hierarchy of the sciences: A Latourian survey of inscription practices. *Social Studies of Science*, 30(1), 73–94. <https://doi.org/10.1177/030631200030001003>
- Valipour, V., Aidinlu, N. A., & Asl, H. D. (2016). Investigating Lexico-grammaticality in Academic Abstracts and Their Full Research Papers from a Diachronic Perspective. *DOAJ (DOAJ: Directory of Open Access Journals)*, 9(19), 178–197. <https://doaj.org/article/8b1b78678b584a24be335aab50b40f53>
- Zuchowski, L. C. (2012). Disentangling complexity from randomness and chaos. *Entropy*, 14(2), 177–212. <https://doi.org/10.3390/e14020177>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.