International Journal of Applied Linguistics

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Do Formal Stance Strategies Reveal Disciplinary Variation in Professional Scientific Writing?

Elizaveta A. Smirnova[1] | Javier Pérez-Guerra[2]

[1]School of Foreign Languages, HSE University, Perm, Russia | [2]Department of English, Universidade de Vigo, Vigo, Spain

Correspondence: Javier Pérez-Guerra (jperez@uvigo.gal)

Funding: Javier Pérez-Guerra is grateful to the Spanish State Research Agency and the European Regional Development Fund (MCIN/AEI/10.13039/501100011033–PID2020-117541GB-I00), and Xunta de Galicia (grant no. ED431C 2021/52) for generous financial support.

Keywords: academic discourse | corpus linguistics | hard sciences | register | soft sciences | stance | discurso académico | postura autoral | registro | ciencias duras | ciencias blandas | lingüística de corpus

ABSTRACT

Stance in academic discourse has been extensively studied, with numerous investigations indicating that its expression varies across disciplines, depending on the authors' intention to either enhance or diminish their voice or presence (e.g., *It seems fairly certain* vs. *This is based on the belief that…*). This paper hypothesises that stance can be viewed as a strictly structural or formal linguistic mechanism in academic discourse, which can optimally determine disciplinary variation. The novelty of this study lies in the hypothesis that formal features of stance alone can identify academic disciplines, without relying on the meaning conveyed by the features. To demonstrate this, this paper focuses on the linguistic expression of stance in hard- and soft-science articles. The corpus of soft and hard scientific writing consists of research articles published in leading peer-reviewed journals in eight disciplines (chemistry, physics, engineering, mathematics, business studies, history, linguistics and political science) and comprises approximately 1.6 million words. The assessment of the realisation of stance in the aforementioned scientific disciplines is carried out by quantifying a range of grammatical (e.g., modal verb groups and embedded complement clauses) and lexical (boosters, hedges, and self-mention expressions) features suggested in the literature. The frequencies of the features are statistically modelled by means of, firstly, a multivariate regression analysis that determines the set of features whose contribution to the hard- versus soft-science variation is significant and, secondly, a clustering technique that groups similar disciplines based on exclusively the frequencies of the significant stance features. Clustering very successfully reveals a neat classification of the eight disciplines under investigation into two major clusters corresponding to the initial categorisation of the writings into the hard- and soft-science types. This suggests that the meaning conveyed by the stance features is dispensable for the purpose of disciplinary categorisation.

ABSTRACT

La 'postura' (o 'actitud') autoral en el discurso académico ha sido estudiada ampliamente, siendo numerosas las investigaciones que sugieren que su expresión varía por disciplina según sea la intención del autor/a aumentar o disminuir su presencia (por ejemplo, en inglés *It seems fairly certain* frente a *This is based on the belief that…*). Este artículo plantea la hipótesis de que la postura autoral puede tratarse de un mecanismo estrictamente estructural o formal en el discurso académico, capaz de determinar de manera óptima la variación entre disciplinas. La novedad de este estudio radicaría así en la hipótesis de que las características formales que marcan postura autoral, por sí solas, podrían identificar disciplinas académicas sin depender del significado asociado a dichas estrategias.

Para demostrar esto, el artículo se centra en la expresión lingüística de la postura autoral en artículos de ciencias 'duras' y 'blandas'. El corpus de textos científicos de ambas categorías científicas consta de artículos de investigación publicados en revistas de impacto con revisión por pares en ocho disciplinas (química, física, ingeniería, matemáticas, estudios empresariales, historia, lingüística y ciencias políticas) con una extensión de aproximadamente 1,6 millones de palabras. La evaluación de la formalización de la postura autoral en dichas disciplinas científicas se ha llevado a cabo cuantificando una serie de rasgos gramaticales (por ejemplo, grupos verbales modales, cláusulas de complemento incrustadas) y léxicos (reforzadores, atenuadores, expresiones de auto-referencia) propuestos en estudios previos.

Las frecuencias de dichos rasgos se han modelado estadísticamente, en primer lugar, mediante un análisis de regresión multivariante que ha determinado el conjunto de rasgos cuya contribución a la variación entre ciencias duras y blandas es significativa y, en segundo lugar, a través de una técnica de agrupamiento o *clustering* que identifica disciplinas similares basándose exclusivamente en las frecuencias de los rasgos significativos de postura autoral. El agrupamiento resultante refleja con gran acierto la clasificación de las ocho disciplinas analizadas en dos grupos principales de ciencias que se corresponden exactamente con la categorización inicial de los textos en ciencias duras y blandas. Se demuestra así que el significado asociado a los rasgos de postura autoral es prescindible a la hora de categorizar disciplinas.

1 | Introduction

Despite the widely assumed aim of research writing to achieve objectiveness and impersonality, previous research has demonstrated that academic prose evinces 'a persuasive endeavour involving interaction between writers and readers' (Hyland 2005, 173). Authors tend to express opinions and attitudes in their texts rather than merely communicate facts (see, e.g., Biber 2006b, 113). The linguistic representation of stance in academic discourse is defined as 'the expression of the speaker's or writer's personal feelings, attitudes, value judgements, or assessments' (Biber et al. 1999, 966). In consequence, stance has been a thoroughly researched topic in the field of academic discourse studies through the examination of linguistic strategies or features of stance. In this respect, it has been taken for granted that the features selected by authors to mark stance are conditioned by the semantics of stance itself (see Section 2.1).

As will be discussed in Section 2, the investigation of disciplinary variation in the expression of stance in professional academic writing is hampered by the fact that disciplines have specific goals, epistemological assumptions and modes of inquiry that might influence the selection of the patterns and strategies that express stance, which has paved the way for discipline-specific studies of stance expression.

This study revisits the way in which stance is materialised in academic writing and undertakes a quantitative analysis of stance features in a representative selection of disciplines. The hypothesis in this investigation is that the expression of stance is, as has been shown in many studies, discipline-specific and can be determined by the distribution of specific linguistic features, irrespective of their semantic profiling. The novelty of this investigation lies in the fact that it evinces that stance features can serve as perfect proxies for the identification of academic disciplines, without resorting to their semantic load. In other words, the objective of our study is not to establish that authors from different academic disciplines employ stance expressions differently. Rather, we aim to demonstrate, through objective, empirical, statistical methodologies, that the linguistic—'formal'—strategies identified in the literature as proxies for stance can effectively distinguish between academic disciplines. Furthermore, our analysis reveals that the association between formal stance features and disciplinary variation is rooted entirely in morphosyntactic or grammatical properties, and can be independent of semantic content.

In detail, this study is carried out on a corpus of research papers, that are seen as 'the key genre of the academy' (see, e.g., Hyland and Jiang 2017, 1), in four soft sciences (business studies, linguistics, history and political science) and four hard sciences (mathematics, engineering, chemistry and physics), published in leading peer-reviewed journals. The goal is twofold: first, to describe the markers of stance employed by professional authors and, second, to test the hypothesis that there is variation in the expression of stance across disciplines. The two research questions (RQs) addressed in this study are:

- RQ1: How is stance expressed in the hard and soft sciences?
- RQ2: Can stance markers serve as proxies for the characterisation of hard and soft sciences, without resorting to semantics?

These research questions will be approached by adopting corpus-linguistic techniques ranging from frequency distributions to more advanced multivariable statistical methods.

This paper is organised as follows. Section 2 outlines the previous works on stance in academic discourse and disciplinary variation. Section 3 describes the data and method of analysis used. Section 4 presents the results of the analysis, which are summarised and discussed in Section 5. Section 6 reports the main conclusions.

2 | Literature Review

The linguistic expression of speakers' or writers' personal feelings and assessments has been termed differently in the literature: 'evaluation' (e.g., Hunston 1994; Hunston and Thompson 2000), 'appraisal' (Martin and White 2005), 'subjectivity' (e.g.,

Fitzmaurice 2004; Ehret and Taboada 2021) or 'stance' (e.g., Biber and Finegan 1989; Conrad and Biber 2000; Hyland 2005), among others. In the linguistic literature, stance has been approached in different ways, ranging from studies that focus on a single text or part of it (see, e.g., Thompson 1993; Tse 2012) to corpus analyses of patterns employed in a number of various texts (e.g., Hunston 1994; Kim and Crosthwaite 2019), as well as from the investigation of a single stance feature (e.g., Hyland and Tse 2005a) to descriptions of a bulk of lexical and grammatical markers of stance (Gray and Biber 2012). This section summarises previous studies on types of stance features recognised in the literature (Section 2.1), on stance in academic genres and specifically on disciplinary variation in the expression of stance (Section 2.2).

2.1 | Types of Stance Features

This study approaches stance by exploring the frequency of an inventory of lexical and grammatical stance features identified by Hyland (2005), Biber and colleagues (mainly, Biber 2006a).

Hyland (2005) posits that a text constructs certain interactions between readers and writers, and suggests that there are two key resources of such academic interaction: the *stance* features that project the writer's propositions and the *engagement* markers that are used to address and involve the reader. In Hyland's view, stance can be materialised by four types of markers: (i) hedges such as *possible* and *may*, which indicate a lack of complete commitment to a proposition (e.g., *Such experiments may not quantitatively represent the amount of embolism that is formed during winter freezing in nature*); (ii) boosters such as *obvious* and *must*, which express the writer's certainty about the topic and his/her alignment with the reader (e.g., *With a few interesting exceptions we obviously do not see a static image as moving*); (iii) attitude markers such as *prefer* and *important*, which reveal the writer's attitude toward propositions and convey affective feelings like agreement or importance (e.g., *two quantities are rather important and, for this reason, the way they were measured is re-explained here*); and (iv) self-mention markers, which involve the presence of the writer through the use of first-person pronouns and possessive adjectives (e.g., *Of course we know there are researchers there, making interpretations and so on, but this is just assumed*). In line with Hyland (2005), Wang and Pramoolsook (2021, 104) approach stance as 'the projection of the writers in the text and the interaction constructed between writers and the readers in terms of the writer-oriented linguistic features'. These authors investigated stance in L2 MA students' abstracts and extended Hyland's (2005) list of stance markers with several genre-specific linguistic units, leading to the list in Appendix 1.

Stance as a dimension of genre variation has previously been studied within Biber's multidimensional analysis. This analytical framework is based on the assumption that register differences are characterised by the systematic use of patterns of co-occurring features, which is rooted in Halliday's (1988) systemic-functional tenets. If particular linguistic features are found to consistently co-occur in a specific text type or register, then there should be 'an underlying functional influence that encourages their use' (Biber 1988, 13). The linguistic features included in each dimension of variation are determined by factor analyses conducted for particular data. Subsequently, the identified factors or dimensions are interpreted by the researcher. The dimensions, as well as the linguistic features with which they are associated, vary in different studies. For example, In Veirano Pinto (2014), the proposed dimensional model of variation in American movies, which includes the dimension 'Expression of stance vs Expression of information', involves *to*- and *that*-complement clauses, *wh*-clauses, pronouns and different types of verbs as features that (positively) contribute to marking stance. In Egbert (2015), devoted to the study of sub-register and discipline variation in published academic writing, the dimension 'Author-centered stance' includes linguistic features such as first-person pronouns, infinitives, *that*-relative clauses controlled by stance nouns and communication verbs. Hardy and Römer (2013), with data from the Michigan Corpus of Upper-Level Student Papers corpus (MICUSP), recognise 'Expression of opinions and mental processes' among their dimensions of disciplinary and genre variation. This dimension is characterised by different types of complement clauses controlled by verbs, predicative adjectives and verbs of communication, among other features.

Gray and Biber (2015, 223) also claim that stance must be investigated through the occurrence of lexico-grammatical features, that is "sets of lexical items that (a) express the various stance meanings and (b) typically occur in these specific grammatical environments". For instance, stance adjectives and nouns governing complement clauses and *of*-phrases (as in (1) are examples of lexico-grammatical markers that are syntactically embedded within larger constituents that provide the context and the shared background readers or hearers need to interpret the writer's/speaker's attitude or evaluation.

1. Wiener's and Bigelow's discovery [of the general importance of feedback loops] is of great interest in the history of Communication. (Gray and Biber 2015, 234)

According to Biber and Finegan (1989), stance comprises evidentiality (the author's attitude toward knowledge, its reliability and the source of information) and affect (the expression of personal feelings and emotions). In this vein, they distinguish six types of adverbial stance markers: (i) HONESTLY adverbials, which mark the manner of speaking (e.g., *I, personally, spoke with my heart*); (ii) GENERALLY adverbials, which deal with generalisation or approximation (e.g., *the anti-communists were in general united only in their anti-communism*); (iii) SURELY adverbials, which express conviction or certainty (e.g., *and, of course, 29-year-old Jerry, to whom…*); (iv) ACTUALLY adverbials, concerned with actuality or emphasis (e.g., *Infact, the Tories made it worse…*); (v) MAYBE adverbials, connected with possibility or hedging (e.g., *Many people feel we have had, perhaps, too much shelter*); and (vi) AMAZINGLY adverbials, used to express the speaker's/ writer's affective attitude toward the content (e.g., *but fortunately they can't be seen by the public*).

Whereas, taken together, the stance features play a significant role in positioning academic arguments within the interactions of disciplinary groups and serve to construct understanding and provide a framework for the interpretation of how authors and audiences establish connections within their disciplinary contexts through written discourse, the meaning conveyed by the features is not uniform. The semantics of the stance features can be approached in different ways. On the one hand, stance can be

determined by the author's attitude toward either their work or the work of others. Thus, Biber and Finegan (1989) classify stance into affect, which can be positive or negative, and evidentiality, including certainty or doubt. To express this attitude, the writer adopts 'a characteristic voice, a particular pattern of syntactic and lexical semantic choices' (Gross and Chesley 2012, 85). Since the evaluation of the statements in a text varies across disciplines (Hyland 2005; see also Section 2.2 in this respect), the selection and use of the stance features will also be affected by disciplinary specificity as well (see, e.g., Hyland and Jiang 2016; Kim and Crosthwaite 2019). On the other hand, stance expressions can be seen as a means for enhancing or diminishing the authors' presence in the text. According to Hyland (2005), writers typically choose whether or not to include explicit references to themselves as authors, which helps establish a specific stance and authorial identity within their discipline.

To give some examples of the semantic categorisation of stance features, the employment of the category of self-mention expressions, as contended by Hyland, highlights the author's presence in the texts (as in (2)). Similarly, the use of hedges (3) and boosters (4) also 'represent an author's explicit intrusion into a text to convey a personal assessment on what is being written about' (Hyland 2012, 145). By contrast, extraposed *to*- and *that*-clauses with expletive *it* as a subject express stance more implicitly, as these constructions do not overtly identify the author's presence (Biber et al. 1999, 977). Complement clauses controlled by adjectives with expletive *it* as a subject and complement clauses controlled by nouns as in (5) and (6), respectively, implicitly suggest that the statements are generally accepted evaluations that go beyond the personal beliefs of the author (Gray and Biber 2015).

2. In addition, *I* expect the external pressures induced by higher advocacy salience to affect the internal dynamics within the diverse coalition favourably. (AJPS-2019-2)
3. It *seems fairly* certain, however, that under Kennedy a negotiated withdrawal was at least an option… (WP-2016-4)
4. *Surely*, these problems do not arise in hybrid inflation models. (EPJ -2019-3)
5. Although it is *useful to* consider conflicting goals and goal prioritization as a function of their importance and of the performance relative to the aspiration level… (AMA-2020-3)
6. This broader notion of perceived obligations towards society is based on the belief *that business firms are an integral part of society*… (JMS-2016-2)

The contributions reviewed in this section have been used to determine the list of features that will be employed in the present study (see Appendix 1).

2.2 | Stance in Academic Registers

As already pointed out in Section 1, although the primary objective of research writing is to achieve objectivity and impersonality, academic prose is not devoid of interaction strategies between writers and readers. Hyland (2004, 156) claims that research into the use of stance markers in research writing in different sciences is likely to facilitate 'the uncovering of the rhetorical and social distinctiveness of disciplinary communities'. In fact, over the last few decades, this has constituted a promising area of research. In this vein, this section focuses on the expression of stance in academic registers, particularly in research articles.

Scholars have analysed stance features in various academic registers, both learner and professional, including doctoral theses (e.g., Thompson 2012; Hyland and Zou 2021; Wu and Paltridge 2021), textbooks (e.g., Biber 2006a, b; Bondi 2012), book reviews (e.g., Salager-Meyer et al. 2012; Zou and Hyland 2022) and research articles (e.g., Hyland 1999; Cheng and Unsworth 2016; Yang 2016). In what follows, we summarise relevant contributions that have explored the linguistic expression of stance in academic registers.

In Biber's (2006a) seminal work, stance features like the ones described in Section 2.1 were explored in academic spoken (classroom teaching, class management talk, office hours, study groups) and written (textbooks, course packs, syllabi, institutional writing) registers. It was concluded that the stance markers tend to be more frequent in the spoken than the written registers, with modal verbs being the most common stance category. Stance adverbs and stance complement clauses were also found to occur more regularly in the spoken registers than in the written registers.

Hyland (2005) analysed the use of hedges, boosters, attitude markers and self-mention expressions (see Section 2.1) as markers of stance in research papers in eight disciplines: mechanical engineering, electrical engineering, marketing, philosophy, sociology, applied linguistics, physics and microbiology. The analysis of the data revealed that hedges were noticeably more frequent in the soft disciplines. Furthermore, the authorial presence is minimal in the hard disciplines, while in the soft sciences, where 'what counts as adequate explanation is less assured, interpretative variation increases and writers must rely to a greater extent on a personal projection into the text', self-mention and attitude markers were found to be more common (Hyland 2005, 188). Hyland and Jiang (2016) studied the distribution of the same stance features over five decades across four disciplines. The authors demonstrated the presence of 'slow changes in traditional knowledge construction practices' with a decrease in overt stance expression in the 'more discursive fields', namely sociology and applied linguistics, as compared to the significant increase in stance features attested by the biology and the electrical engineering texts (Hyland and Jiang 2016, 269). Hyland and Jiang (2017, 21), claimed that even though the popularity of evaluative *that*-clauses declined in the period 1965–2015 across all the disciplines that they considered, the increased frequency of this strategy per paper 'suggests that it remains a significant rhetorical option for authors'. To round off the review of studies engaging with the connection between stance and academic register(s), Kim and Crosthwaite (2019) investigated the use of *that*-complement clauses in a corpus of research articles in business and medicine. Their analysis showed that *that*-complement clauses are frequently employed by authors to project their own and previous findings. Besides, it was found that the most common type of *that*-clause in both disciplines is controlled by verbs, and it is used to convey an epistemic stance. As for disciplinary differences, it was demonstrated that the use of evaluative *that*-clauses is significantly higher in business than in medical papers. Xie et al.

(2024) conducted a diachronic analysis of the stance features proposed by Hyland (2005) in Chinese MA theses and research articles in linguistics. A key finding is that Chinese students use fewer stance markers in their writing, as compared to professional writers, over the considered time span. This suggests that students may be hesitant to express their own opinions in their writing, possibly due to factors like their status as novice writers, lack of confidence in their knowledge, the audience they are writing for, or strict supervision and instruction in EAP writing. Hyland and Zou (2021) investigated stance in spoken academic discourse. They studied the use of Hyland's (2005) features in doctoral students' presentations in social and hard sciences. Their analysis showed that, overall, the stance markers are more frequently used in the hard sciences, where they are mostly realised by hedges and boosters. These markers indicate the speaker's opinion on the trustworthiness of a statement, either by questioning or confirming it, which is commonly required in the hard sciences. These results were confirmed by Qiu and Jiang (2021), who also investigated interaction features in presentations. They proved that hard sciences employ stance expressions, especially attitude markers, to a larger extent than other disciplines. In their words, attitude markers are used in hard-science disciplines to 'highlight the importance and surprise value of research findings' (Qiu and Jiang 2021, 10). Chan (2015) investigated the lexico-grammatical stance devices in Biber (2006b) in dissertation acknowledgements in various disciplines. The author identified significant differences in the use of the stance features between soft and hard sciences. To illustrate this, adverbs and complement clauses were found to be more frequently employed in the soft disciplines, while modal verbs are more common in the hard disciplines.

The studies summarised here and, in particular, the controversial results arising from the reviewed literature pave the way for a fine-grained exploration of the linguistic strategies of the expression of stance across the two broad groups of academic disciplines, hard and soft, that constitute the empirical basis of this research, under the hypothesis that the stance strategies employed diverge in hard and soft sciences as a consequence of the different degrees of subjectivity of such discourses.

3 | Data and Methodology

This study analyses professional academic writings retrieved from a corpus of research articles published in high-profile peer-reviewed journals. The data comprise papers in four hard disciplines and four soft disciplines that represent a broad cross-section of academic discourse. The earliest attested use of the label 'hard science' dates back to the nineteenth century when it was employed by Winkworth et al. (1858) in the *Journal of the Society of Arts*. Storer (1967) employed the labels 'hard' and 'soft' to assess scientific disciplines in terms of perceived methodological rigour, precision and objectivity. In summary, the hard disciplines encompass applied, empirical, experimental and natural sciences (e.g., physics, biology, astronomy, mathematics, technology), while the soft disciplines include the social sciences and humanities (e.g., psychology, sociology, political science, philosophy, linguistics, literature, history). We contend that the hard/soft division can serve as a useful shorthand when attempting to describe the diversity of academic discourse (see Dang 2018), provided that comparing the two groups is taken with caution to avoid overlooking the diversity and dynamism of various disciplines (Thompson and Hunston 2019).

The hard-science subcorpus used in this study consists of articles in chemistry, physics, mathematics and engineering, and the soft-science subcorpus includes papers in business studies, history, linguistics and political science research articles. The choice of disciplines is based on previous research in which they were contrasted (see, e.g., Becher and Trowler 2001; Huang and Chang 2008; Gray 2015). All the articles were published in leading academic journals, indexed within the first quartile of Scopus, spanning the years 2016–2020. The papers were randomly selected and manually retrieved from the publishers' websites. Within each discipline, we sampled a number of articles to ensure that the subcorpora are balanced as regards token numbers. The texts underwent formatting for subsequent textual analysis, involving the removal of elements such as tables, formulas, graphs, charts, metadata and reference lists from the documents. Detailed information about the corpus is given in Figure 1.

This investigation undertakes the exploration of lexical and grammatical strategies denoting subjectivity taken from previous research on stance in academic discourse, namely Biber's (2006a) and Hyland's (2005) lists of stance expressions (see Section 2.1), together with other features suggested in Wang and Pramoolsook (2021). The rationale behind the investigation of these stance features is twofold: First, as mentioned above, the linguistic units were previously found to be common in academic texts; therefore, they can be expected to be frequently used in our corpus. Second, complementing Biber's list with Hyland's and Wang and Pramoolsook's interactional metadiscourse items may help to get a more comprehensive picture of the expression of stance in the discourse of research articles.

The linguistic features under consideration are (see Appendix 1 for the full list of linguistic items considered in this study):

- modal and semi-modal verbs of three categories (possibility/permission/ability, e.g., *can*; necessity/obligation, e.g., *must*; prediction/volition, e.g., *will*)
- stance adverbs of two types: epistemic (certainty, e.g., *certainly*; likelihood, e.g., *possibly*), attitude (e.g., *hopefully*) and style (e.g., *mainly*)
- stance complement clauses:
  - controlled by adjectives, of two types: *that*-clauses (e.g., *it is apparent that we should mention this*) and *to*-clauses (e.g., *it is necessary to mention this*)
  - controlled by verbs, of two types: *that*-clauses (e.g., *the research showed that this must be analysed in detail*) and *to*-clauses (e.g., *we decided to analyse this in detail*)
  - controlled by nouns, of two types: *that*-clauses (e.g., *it's a fact that we have to start right now*) and *to*-clauses (e.g., *we had an intention to start right now*)
- hedges (e.g., *claim*, *from this perspective*)
- boosters (e.g., *evidently*, *prove*)
- attitude markers (e.g., *dramatic*, *easily*)
- self-mention expressions (e.g., *we*, *the authors*)

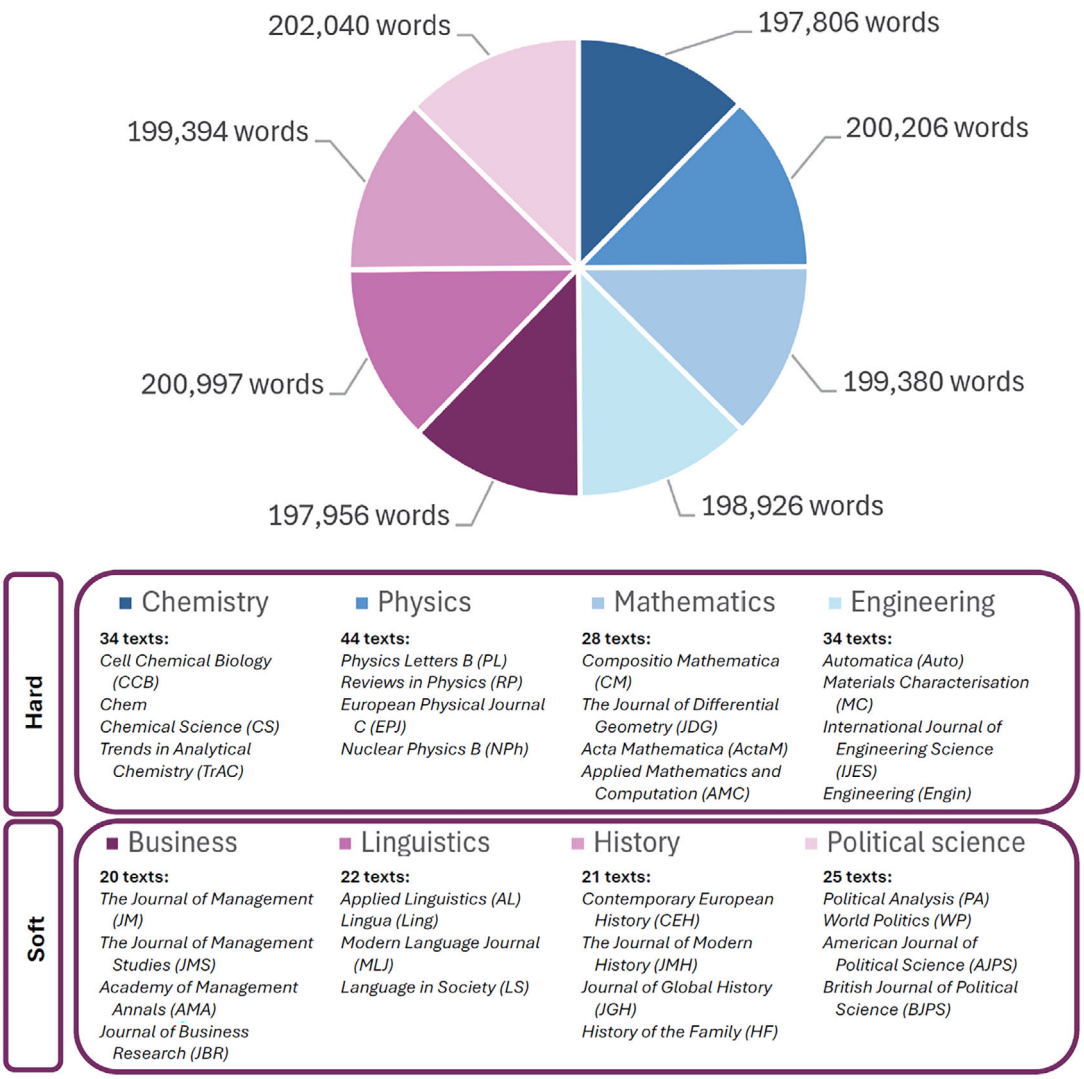

FIGURE 1 | Corpus. [Color figure can be viewed at wileyonlinelibrary.com]

The corpus was processed with AntConc (Anthony 2014), which eased the quantitative analysis of the stance features listed above. The search inventory included the items given in Appendix 1 that were successively searched for with the Advanced Search tool of AntConc. The data analysis necessitated thorough manual pruning and meticulous interpretation of the examples; for instance, the word *even*, on the list of attitude markers in Hyland (2005), is frequently used as an adjective referring to numbers that can be divided exactly by 2 (in (7)) in, for example, maths papers, and such instances had to be discarded from the analysis.

7. The computation for n odd is more straightforward; for n *even*, the analogous results are somewhat more complicated, but not essentially different. (CM-2016-2)

The coding of the stance features was carried out by two trained research assistants under the supervision of the authors. Training included practice trials with the coding procedure and discussion with the supervisors to ensure intercoder reliability and consistency of the results. In a few instances of disagreement, the authors made the final decision.

4 | Analysis of the Data

This section describes the trends manifested by the distribution of the stance features across disciplines. Also, disciplinary variation has been statistically modelled to assess the contribution of stance features to the hard/soft distinction. The aim is to discern similarities and differences in stance among the hard and soft disciplines.

4.1 | The Distribution of Stance Features Across Disciplines

The first stage involves the analysis of the distribution of the stance features across hard/soft disciplines. The frequencies for the categories of features per disciplinary family are displayed in Figure 2 (see Appendix 2 for the normalised frequencies of the stance markers per 100,000 words).

Figure 2 shows that the most common stance features in the corpus are modal verbs (in (8)), complement clauses controlled by verbs (in (9)), hedges (in (10)) and self-mention expressions (as

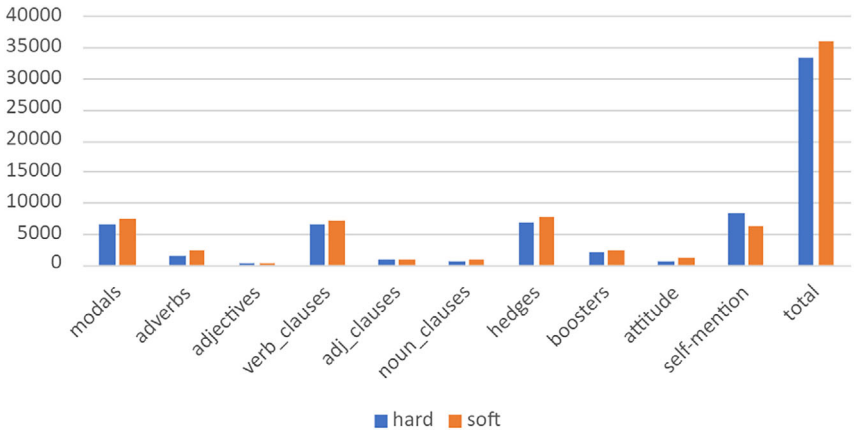

FIGURE 2 | Raw frequencies of stance features. [Color figure can be viewed at wileyonlinelibrary.com]

in (11)). The most frequent marker is self-mention, significantly more common in hard sciences ($\chi^2 = 414.98$, $p < 0.0001$).[1] By contrast, modals, verb-controlled complement clauses and hedges are significantly more frequent in hard-science writing ($\chi^2 = 55.32$, $p < 0.0001$ for modal verbs, $\chi^2 = 25.45$, $p < 0.0001$ for complement clauses controlled be verbs, and $\chi^2 = 50.06$, $p < 0.0001$ for hedges).

8. Therefore biases in the metabolomics data *should* be quantified, potentially corrected, and always reported. (Trac-2018-1)
9. …still, 50 inclusions *seem to be sufficient* to have numerically isotropic behaviour of the composite. (IJES-2020-3)
10. In most real-world applications, the actor of interest will *probably* be an incumbent party or coalition… (PA-2020-2)
11. *We* performed similar radioactive labeling studies in E. coli, a bacterium that normally does not allow exogenous fatty acids to enter the fatty acid synthetic pathway. (CCB-2016-1)

Although the overall distribution of the general lexical and grammatical types is fairly similar in the two disciplinary families, stance is explicitly expressed in the soft-science writings to a greater extent than in the hard disciplines ($\chi^2 = 79.75$, $p < 0.0001$), with the exception of a few (more fine-grained) features that are more common in the hard sciences, namely, complement *that*-clauses controlled by epistemic verbs (examples (12) and (13)) ($\chi^2 = 129.4$, $p < 0.0001$ and $\chi^2 = 48.47$, $p < 0.0001$, respectively), *to*-clauses controlled by adjectives of difficulty (in (14)) ($\chi^2 = 18.07$, $p< 0.0001$) and, to a lesser extent, by verbs of cognition (in (15)) ($\chi^2 = 6.2$, $p = 0.0124$), and, as already reported, expressions of self-mention.

12. *This means that*, even at the lowest limit, a 10% gain in the transfer rate between the DBV and PCB pigments can be realized. (Chem-2016-1)
13. Recently, *it was predicted that* knowing that a quantum state is sparse can facilitate characterization of three-photon quantum states just by measuring two-photon coincidences. (RP-2016-5)
14. This is not *possible to* see directly if one starts with f (T) gravity (PL-2017-2).
15. This implies that when Σ is *known to have* finite index, then we may also compute the index by taking the limit of extrinsic balls. (JDG-2016-2)

4.2 | Statistical Modelling of the Stance Features

This section describes the statistical analysis of the data. Building on the hard/soft difference as regards the frequency of stance strategies (see Section 4.1) and the findings reported in the relevant literature (see Section 2.2), we elaborate on the premise that scientific disciplines can be categorised into hard/soft according to exclusively the pervasiveness of the linguistic strategies expressing stance without resorting to the semantic contribution of such strategies.

The first step involves the identification of the features that contribute to the hard/soft distinction. To do that, we have developed a multivariable model based on a pool of independent stance predictors or features that have been claimed to have a bearing on the characterisation of hard and soft scientific writings, according to the literature reviewed in Section 2. Owing to the scarcity of examples in some of the subcategories, the number of predictors described in Section 2.1 and in Appendix 1 that could have a determining effect on the dependent variable hard/soft had to be reduced to allow for the implementation of a multivariate model. Specifically, given the number of tokens in the smaller level of the dependent variable hard/soft (88 occurrences of stance features in the soft-science subcorpus), we could not consider more than nine independent predictors. The statistical optimisation of the database led to the final list of variables of hard/soft variation in Table 1.[2] We want to highlight that none of the predictors encode the semantic effects of the formal stance features.

To ascertain the relative weights of numeric predictors within a multivariate model, we employ two statistical techniques: regression analysis and random forests—derived effects plots will visualise the association between the stance features that significantly characterise the hard/soft variation. Finally, the

TABLE 1 | Summary of the database.

Variable	Levels	Comments
hardsoft	Hard Soft	
discipline	Business Chemistry Engineering History Linguistics mathematics Physics Politicalsc(ience)	
modal (modal auxiliaries)	Numeric	Includes possibility, necessity and prediction modals.
adv (adverbs)	Numeric	Includes epistemic-certainty, epistemic-likelihood, attitude and style adverbs.
adj_compl (complement clauses controlled by adjectives)	Numeric	Includes certainty, likelihood, attitude and evaluation *that*-clause complements, and certainty, likelihood, attitude, evaluation, ability and difficulty *to*-clause complements.
v_compl (complement clauses controlled by verbs)	Numeric	Includes certainty, likelihood, attitude and speech *that*-clause complements, and probability, cognition, desire, causation and speech *to*-clause complements.
n_compl (complement clauses controlled by nouns)	Numeric	Includes certainty, likelihood, attitude and communication *that*-clause complements, and *to*-clause complements.
hedge	Numeric	
booster	Numeric	
attitude	Numeric	
selfmention	Numeric	

clustering of the disciplines by solely relying on the frequencies of the stance predictors in the individual texts will employed to corroborate whether the variables encoding formal (non-semantic) stance features selected in this study can serve as proxies for the hard/soft distinction.

First, a binomial regression analysis was employed on the dataset: a fixed-effects model using the function 'glm' from the 'stats' package (R Core Development Team 2024) and 'lrm' from the 'rms' package (Harrell Jr and Frank 2022) in R (R Core Team 2023).[3] No severe collinearity between predictors was revealed by their 'vif' values (variance inflation factor, 'car' package; Fox and Weisberg 2019), in all cases lower than 3.8.[4] To find the model that best describes the variation in the data, we used backward stepwise elimination[5] (function 'step' from the 'MASS' package; Venables and Ripley 2002), which revealed that a reduced model with only the factors 'adv' (stance adverbs), 'adj_compl' (complement clauses governed by adjectives), 'v_compl' (complement clauses governed by verbs), 'n_compl' (complement clauses governed by nouns) and 'attitude' markers explains the variation without any statistically significant loss, as determined by AIC (Akaike Information Criterion) comparison (full model's AIC = 164.22 vs. reduced model's AIC = 159.8; ANOVA p = 0.4658). The evaluation of the reduced model confirmed its remarkable discrimination and classification accuracy (C-index = 0.928) and explanatory power (Nagelkerke $R^2 = 0.674$). Table 2 reports the output of the model, with predicted odds for soft disciplines.

Second, the regression model outlined in the preceding paragraph to evaluate the significance of the predictors' contribution to the overall variation has been given support by random forests. Random forests (utilising the 'cforest' function from the 'party' package; Hothorn et al. 2006), initially introduced to linguistic analysis by Tagliamonte and Baayen (2012), were employed to rank the fixed predictors based on their impact on explaining the variation. Figure 3 illustrates the scale of the variables' conditional importance after applying random forests, with the C-index of the random forests (0.963) surpassing to a slight extent, as anticipated, to that of the regression model. The figure reflects the significant impact of the fixed predictors 'adv', 'n_compl', 'adj_compl' and 'attitude', and the more minor contribution of 'v_compl' to the model.

Thirdly, to gain a more comprehensive understanding of the influence specifically wielded by the most significant predictors on the variation as identified by the regression analysis and random forests, Figure 4 displays the corresponding effects plots (function 'predictorEffects' from the 'effects' package; Fox and Weisberg 2018). The graphs plot the predicted probabilities for the hard versus soft variation in each case: the higher the probability of occurrence in soft-science writings, the closer to the topmost part of the graph, a higher incidence in the hard disciplines being reflected in the bottom part of the graph.

The effect plots of the independent variables in Figure 4 demonstrate that the occurrence of stance adverbs, complement clauses

TABLE 2 | Output of the regression model (significance codes: 0 '***'; 0.001 '**'; 0.01 '*'; 0.05 '.'; 0.1 ' ').

Coefficients	Estimate std.	Error	z value	Pr(>|z|)	
(Intercept)	−4.28402	0.56758	−7.548	4.43×10^{-14}	***
adv	0.11416	0.02487	4.590	4.44×10^{-6}	***
adj_compl	0.07225	0.03049	2.370	0.017792	*
v_compl	−0.03094	0.01334	−2.320	0.020338	*
n_compl	0.22026	0.05957	3.698	0.000218	***
attitude	0.09206	0.04234	2.174	0.029699	*

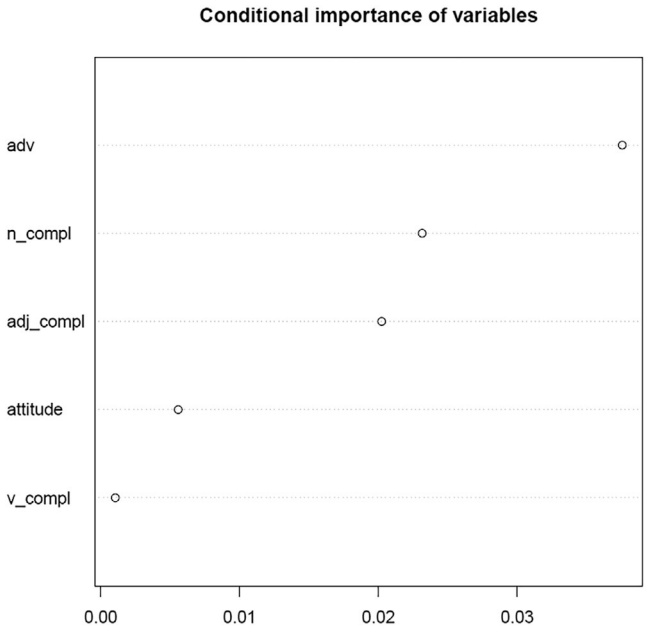

FIGURE 3 | Dot chart of conditional variable importance.

controlled by adjectives and nouns and attitude markers is more salient in the soft sciences, while complement clauses controlled by verbs display a stronger association with the hard sciences. This list of significant features is based on a statistical model that takes into account the contribution of *all* the stance metrics to the hard/soft variation rather than the relevance of individual stance features. This explains the differences between this list of significant features and the one reported in Section 4.1, determined by the individual frequencies of the stance strategies attested in the hard- and the soft-science subcorpora.

As an interim summary, we investigated the formal mechanisms of stance listed in Table 1 in our corpus of hard- and soft-science writings. To determine the inventory of the features that strongly contribute to the hard/soft distinction, which constitutes one of the research questions in this study, we built a statistical model that allowed us to discard a number of predictors whose relevance to the hard/soft variation was not significant. As a result, the statistical methods determined that the frequencies of the features stance adverbs, complement clauses governed by adjectives, complement clauses governed by verbs, complement clauses governed by nouns and attitude markers could confidently predict the (hard/soft) subcorpus to which the examples belong.

With the objective of corroborating the previous results, which did not take into account the specific levelling of the dependent variable hard/soft into the eight disciplines, we have conducted an analysis of stance strategies across the eight 'discipline' levels using a hierarchical agglomerative clustering algorithm[6] ('hclust' function, 'ward.D2' method, 'pvclust' package; Suzuki et al. 2019) based on the five variables entering the reduced regression model. This clustering method has served to identify subgroups of disciplines in light of the pervasiveness of the stance mechanisms. Figure 5 displays a dendrogram of the clustering of the varieties.

The statistical technique organised the eight disciplines into two statistically optimal clusters, depicted by the boxes in Figure 5. The stability and alignment of these clusters with the data were assessed using the 'pvclust' function from the 'pvclust' package. This function employs multiscale bootstrap resampling to compute Approximately Unbiased (AU) p values for each cluster (in red). A higher AU p value indicates greater statistical significance for the cluster, with proximity to 1 indicating outstanding significance.

5 | Results and Discussion

As regards the first research question 'How is stance expressed in the hard and soft sciences?', the empirical techniques and explorations described in Section 4 revealed the following trends revealed by the frequency and distribution of stance expressions in hard/soft scientific writings. To start with, a preliminary analysis of the data in Section 4.1 unveiled significant differences in the overall frequency of stance markers between hard and soft sciences. Subsequently, the statistical models in Section 4.2 demonstrated the following associations between stance features and disciplines. The regression analysis, the effects plots and random forests determined that the soft-science texts contain more stance adverbs, complement clauses controlled by adjectives and nouns, and attitude markers. On the other hand, complement clauses controlled by verbs are common in the hard disciplines.

First, all types of stance adverbs were found to be significantly more frequent in the soft-science writings, the greatest difference affecting epistemic adverbs expressing likelihood, which are more prevalent, for example, in history papers, where the adverbs are used to express authorial speculations on some historic events, as in (16):

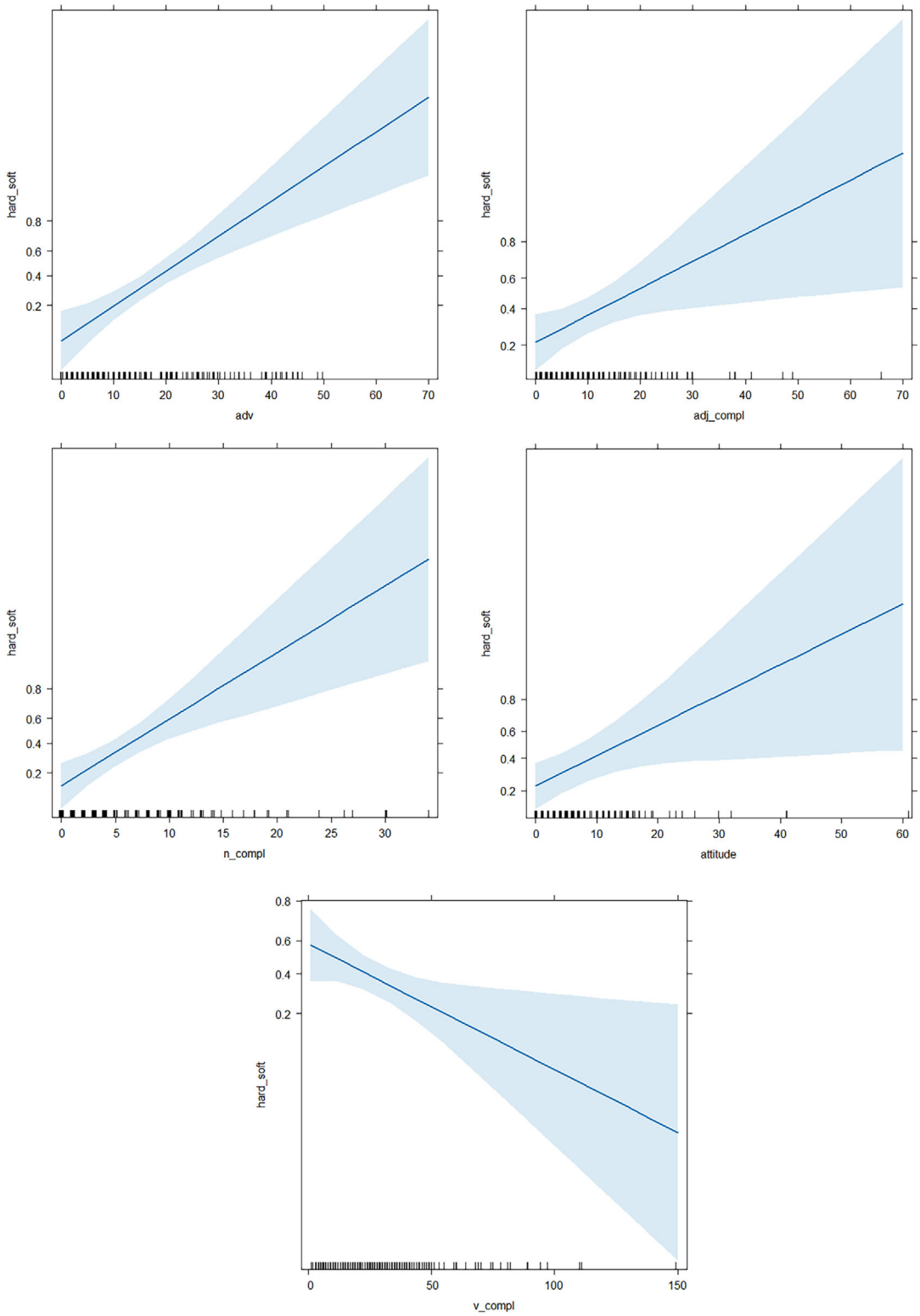

FIGURE 4 | Effects plots. [Color figure can be viewed at wileyonlinelibrary.com]

16. Wealthy parents may have chosen to inoculate their children, and this *probably* accounted for much of the mortality advantage associated with wealth in the period 1752–94. However, this advantage was *probably* very substantially diminished with the very widespread acceptance of and access to vaccination after 1800. (HF-2019-1)

Adverbs of likelihood, which were previously termed as MAYBE adverbials by Biber and Finegan (1988), represent a 'cautious style' common for academic prose, especially in situations when 'the evidence, underlying the assertions is not certain' (Biber and Finegan 1988, 24–25), which is often the case in research in history.

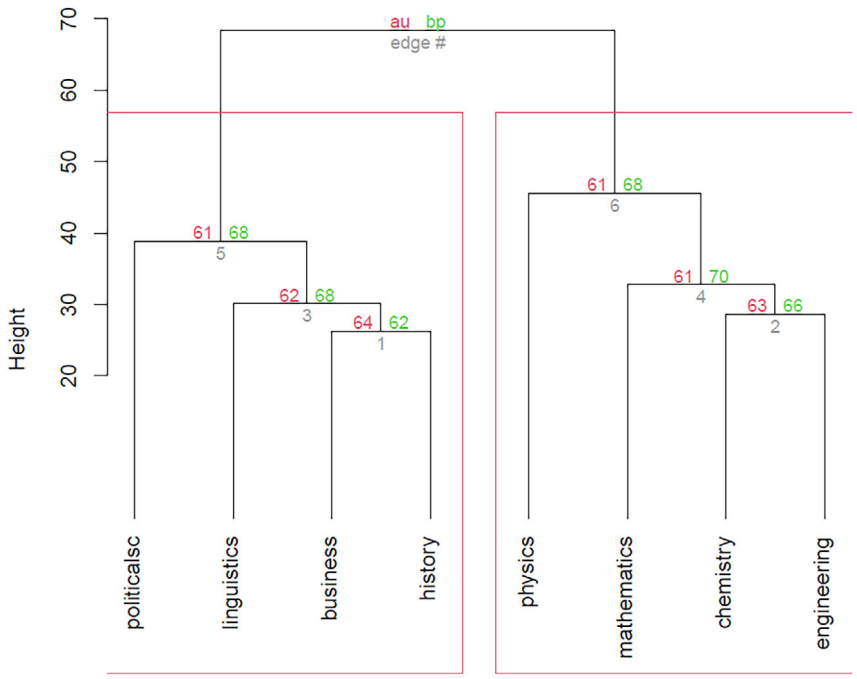

FIGURE 5 | Clustering of disciplines. [Color figure can be viewed at wileyonlinelibrary.com]

Adverbs of style were found to be especially common in linguistics where they are most typically used to 'indicate that a statement describes the usual case, rather than an invariable fact' (Biber 2006a, 105) (17):

17. Another crucial difference with respect to the S-framed pattern is that the goal phrase *typically* lacks an unambiguously directional preposition. This is not surprising given that V-languages *generally* do not have path elements. (Ling-2020-1)

Second, as for complement clauses governed by adjectives, the most striking difference is evinced by *to*-clauses controlled by adjectives expressing likelihood. This stance strategy is not attested in the mathematics papers and is infrequent in the other hard sciences, whilst it is common in the soft sciences, particularly in political science papers, where the *to*-clauses are often employed to comment on quantitative analyses or measures of data or facts (e.g., *twice as likely to…*, *three times to…*), as in (18)—see Section 4.1. As mentioned by Gray and Biber (2015, 230), such structures lack 'explicit attribution to the writer'. In other words, they implicitly suggest that the information is generally accepted, so the reader should accept it as well.

18. We show that US militarized interstate disputes (mIds) that have occurred under Southern presidents *have been twice as likely to* involve the use of force, have lasted on average twice as long, and *have been three times as likely to* be won by the United States. These differences are robust to matching presidencies on potential confounders and *are unlikely to* be the result of selection effects or of Southern presidents' other attributes. (WP-2016-4)

Another type of *to*-complement clauses controlled by adjectives that is significantly more common in the soft sciences is that governed by adjectives of ability. These structures are especially frequent in articles in history and business studies where they are mostly used in case-studies (19) and descriptions of historical events (20).

19. Divided by racial and ideological faultlines (features of identity-based subteams), the overall team *was unable to* communicate or make progress toward its task, leading the course instructors to transfer Ken to another group… (JM-2016-1)

20. No longer *was Finland willing to* limit its policy to safeguarding its most vital export interests; the Finns were now interested in broader benefits of integration. (CEH-2016-2)

To-clauses controlled by nouns are also characteristic of soft-science writings, especially in political science. As exemplified in (21), the nouns governing such infinite complement clauses in this discipline, for example, *threat*, *responsibility*, *decision* or *obligation*, express key concepts on which political scientists focus in their research.

21. When in 1961 Soviet premier Nikita Khrushchev renewed his *threat to cut off* access to West Berlin, Kennedy, fearing the damage to US credibility, refused to back down, and Khrushchev yielded instead. Even more dramatically, Kennedy went 'eyeball to eyeball' with Khrushchev over the latter's *decision to install* nuclear weapons in Cuba despite Kennedy's declaration that he would not tolerate offensive weapons. (WP-2016-4)

Third, political science is also leading the use of attitude markers, characteristic of soft-science writing. The most popular attitude markers employed by the soft-science authors are *interesting/interestingly* and *important/importantly* (accounting

for almost 50%of the occurrences). These features are utilised for commenting on the works of others (see 22), identifying the research gap (23), or describing findings (24).

22. *Interestingly*, Baldwin also finds that voters tend to cast their ballots for political candidates endorsed by chiefs because they infer, *correctly*, that politicians with connections to chiefs will provide higher levels of public goods. (WP-2016-1)
23. something that has long been noted as *important* in corpus enquiry, but never fully embraced or examined owing to the text-based nature of current corpus resources. (AL-2016-2)
24. The current study established that the use of TERRS was the most *important* predictor of reading self-efficacy for beginner learners of French at the start of secondary school, even when reading proficiency and LK were also considered. Three clusters of learners with distinctive profiles emerged, in which TERRS was an *important* distinguishing factor. (MLJ-2020-1)

The extensive use of attitude markers in the soft sciences was previously described by Hyland (2005), who claimed that in those disciplines, due to a potential lack of assured adequate explanation, the weight of the researcher's personal interpretation of information is greater than in hard sciences.

By contrast, the empirical techniques revealed that complement clauses controlled by verbs are more strongly associated with the linguistic repertoire of hard-science stance mechanisms. For instance, *that*-clauses governed by verbs expressing certainty, which constitute the most common construction of this type in the two subcorpora, are significantly more frequent in the hard-science texts, specifically in mathematics, where they are used to comment on calculations—see (25).

25. We shall *prove that* x(0) satisfies the condition (3.21). It is easy *to prove that* conditions (1.5) can be formulated in the following equivalent form where… (AMC-2019-3)

This finding is in line with Hyland and Tse (2005b)*, who found that expressions of certainty, especially controlled by reporting verbs, are noticeably more common in hard sciences (biology, computer science, electronic engineering) than in the humanities. The same authors (Hyland and Tse 2005a, 129) suggest that in academic discourse subordinate clauses can primarily fulfil the following functions: serve to express 'the authors' evaluation of their own findings, their evaluation of previous studies, their statement of research goals, and evaluation of the methods, models, and theories'. According to Kim and Crosthwaite (2019), in sentences with *that*-complement clauses controlled by verbs, the subject and the verb contain both the source of the evaluation and an evaluative stance, while the part after *that* comprises the entity that is evaluated by the writer. The imperatives *suppose/assume* used in formulas make up more than 50% of the instances with *that*-complement clauses controlled by likelihood verbs in the hard sciences (examples (26) and (27), respectively), which instantiate a way to engage with the readers by appealing to collective knowledge in the field and encouraging to follow the authors' ideas.

26. *Suppose that* C(0) is properly aligned and *suppose* for the sake of contradiction *that* there exists Y ⧹xE2^^ A(C(0)) ⧹xE2^⧹xA9 B1/4(0) for which (Rl+k ⧹xC3– {Y}) ⧹xE2^⧹xA9 D(V) contains more than one point. (JDG-2017-3)
27. *Let us assume* that initially at the shell was at rest, then it started to move till and then stopped. (PL-2017-3)

The wide use of *that*-complement clauses in the hard sciences can partly explain the high frequency of self-mentions in the hard-science writings, a finding that is at odds with Hyland's (2005, 181) conclusion that 'in the sciences it is common for writers to downplay their personal role to highlight the phenomena under study, the replicability of research activities, and the generality of the findings, subordinating their own voice to that of unmediated nature'. Mathematics writings were found to employ personal pronouns *I* and *we* more frequently than any other sciences, which contradicts Burton and Morgan's (2000, 435) claim that in this discipline 'the apparent absence of the author from the text fits with positivist epistemologies in which the mathematician's role is subordinate to that of the mathematics itself'. Similarly, Hyland (2005, 188) contends that since hard science writing is highly formalised and tightly structured, the writers tend to 'minimise their presence in their texts' in the hard disciplines. That Hyland's statement is not corroborated by our empirical findings (see Section 4.1) can be explained by the frequent detailed descriptions of calculations in the hard-science articles utilising personal pronouns in our data, particularly in mathematics (see (28)) and physics (29).

28. As s approaches zero from the left, *we* let the length of the finite gradient flow line along N go to zero. In particular, at s = 0, *we* obtain the moduli space shown in Figure 11 which has three disc components meeting at two boundary nodes. (ActaM-2018-2)
29. If *we* relax the requirement of parity preservation, *we* have two new quadratic parity violating invariants [8] which are … *We* can then naturally consider a straightforward generalization of the gravity Lagrangian in the following way… (PL-2017-2)

With respect to the second research question 'Can stance markers serve as proxies for the characterisation of hard and soft sciences?', the statistical analyses of the data identified the stance features that served as strong predictors of the hard/soft variation. Also, by relying exclusively on the frequencies of the most significant stance strategies in the texts, the eight hard/soft disciplines under scrutiny were clustered into two statistically optimal groups that perfectly align with the two broad categories of disciplines being examined, namely hard and soft sciences. This demonstrates that the formal expression of stance can be used to characterise and categorise academic disciplines without the need to consider the meaning conveyed by the stance strategies.

6 | Conclusion

This study has looked at the expression of stance in professional academic writing by analysing a number of grammatical and lexical stance markers in a corpus of research papers in four hard and four soft sciences. As regards the first research question 'How

is stance expressed in the hard and soft sciences?', it has been shown that, in general, stance expressions are more frequently attested in soft sciences, and that the markers that best describe the variation between the hard and soft disciplines are: stance adverbs, complement clauses controlled by adjectives, complement clauses controlled by nouns, complement clauses controlled by verbs and attitude markers. All these stance features, except complement clauses controlled by verbs, are more frequent in soft-science texts. In response to the second research question 'Can stance markers serve as proxies for the characterisation of hard and soft sciences?', the statistical modelling of the data has unveiled two optimal clusters that do coincide with the two categories of disciplines under analysis, that is, hard and soft. In other words, the cluster division of the research articles, based solely on the frequencies of the formal types of stance features attested in the texts, corresponds neatly to the classification of the disciplines into hard (mathematics, engineering, chemistry, physics) and soft (business, linguistics, history, political science). This result effectively demonstrates that the (formal) expression of stance can serve as a valid basis for differentiating these two disciplinary families, which constitutes a significant contribution of this study.

In a nutshell, our study has not focused on proving that stance expressions vary across academic disciplines, which has already been demonstrated in prior studies. Instead, we have aimed to show, using rigorous empirical methods, that the linguistic 'formal' strategies commonly regarded as indicators of stance can reliably differentiate between academic disciplines. This leads to the acceptance of a continuum spanning hard and soft academic disciplines. Moreover, our findings have demonstrated that the link between formal stance features and disciplinary differences is grounded entirely in morphosyntactic or grammatical structures, without necessarily relying on semantic interpretation. To our knowledge, this perspective had not been explored in previous research.

In terms of the limitations of the present research, it seems in order to recognise that the twofold classification of sciences into hard and soft may not capture all disciplinary nuances and that a more comprehensive understanding of stance in academic writing will benefit from the exploration of a more fine-grained inventory of disciplines and genres.

To conclude, the analysis of disciplinary variation in the expression of stance has multiple potential applications. First, it can aid in the development of genre-specific language resources for ESP (English for specific purposes) and EAP (English for academic purposes) courses. Understanding how different disciplines express stances can help learners to write more effectively in their respective fields. Second, it can contribute to the development of automated tools for text analysis and evaluation. In this respect, we have in mind the various types of software that can check the grammar and style of a text (e.g., Grammarly (https://www.grammarly.com), Ginger (https://www.gingersoftware.com/), Language Tool (https://languagetool.org/ru)), detect rhetorical moves in a text (Pendar and Cotos 2008) and even provide feedback about errors (see, for instance, Napolitano and Stent 2009; Dreschler et al. 2019). To the best of our knowledge, to date, the focus of these tools is not on discursive strategies, so tailoring the software so that it can identify, revise and/or suggest the employment of, for example, discipline-specific stance features in texts can assist researchers in improving their academic writings—for example, after inputting their text and specifying their research field, authors could receive detailed feedback on the frequency or productivity of stance features in their manuscript and their alignment with disciplinary conventions, along with suggestions for alternative structures. Additionally, stakeholders in academic publishing, such as journal editors, can use this analysis to get a better comprehension of characteristic features of different disciplines, which might be used for developing guidelines for authors, ensuring a more coherent and accessible academic discourse.

Acknowledgments

We thank Olesya Kolosova and Nuriiat Omarova for their help with data processing.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1111/ijal.12694.

Endnotes

[1] The significance of frequency differences was determined through chi-squared tests and corresponding p values. In cases where values were ≤ 5, the Fisher exact test was employed for p value calculation. The convention for statistical significance was established as follows: '***' when $p \leq 0.001$, '**' when $p \leq 0.01$, and '*' when the significance of the variation is indicated by a p value ≤ 0.05.

[2] The Kolmogorov–Smirnov (KS) test was used to conclude that our data follow a normal distribution ($p = 0.797$).

[3] data_glm <- glm(hardsoft ~ modal + adv + adj_compl + v_compl + n_compl + hedge + booster + attitude + selfmention, data = data, family = binomial).

[4] Collinearity in regression analysis detects highly correlated predictors that provide redundant information. This implies that changes in one predictor are mirrored by changes in another, so their independent contribution to the model is hard to distinguish.

[5] glm_step <- step(data_glm, direction = "backward").

[6] The analysis employed a behavioral profiles approach, wherein the data are represented as vectors indicating proportions of each level for every variable (utilising the 'bp' function from the 'pvclust' package). Numerical differences between vectors are quantified as 'distances' (using the 'dist' function with the 'canberra' method, from the 'pvclust' package), which determine the grouping of varieties into clusters. The optimal number of clusters was determined using the 'silhouette' function from the 'cluster' package (Maechler et al. 2019). These clusters are depicted as tree leaves or branches in dendrograms (Levshina 2015, 316), where the most similar clusters (those with the smallest 'distance') are merged together.

References

Anthony, L. 2014. *AntConc*, version 3.4.4. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/antconc/.

Becher, T., and P. Trowler. 2001. *Academic Tribes and Territories*. New York, US: McGraw-Hill Education.

Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge, UK: Cambridge University Press.

Biber, D. 2006a. *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam, NL: John Benjamins.

Biber, D. 2006b. "Stance in Spoken and Written University Registers." *Journal of English for Academic Purposes* 5, no. 2: 97–116. https://doi.org/10.1016/j.jeap.2006.05.001.

Biber, D., and E. Finegan. 1988. "Adverbial Stance Types in English." *Discourse Processes* 11, no. 1: 1–34. https://doi.org/10.1080/01638538809544689.

Biber, D., and E. Finegan. 1989. "Styles of Stance in English: Lexical and Grammatical Marking of Evidentiality and Affect." *Text* 9, no. 1: 93–124. https://doi.org/10.1515/text.1.1989.9.1.93.

Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan, and R. Quirk. 1999. *Longman Grammar of Spoken and Written English*. London, UK: Longman.

Bondi, M. 2012. "Voice in Textbooks: Between Exposition and Argument." In *Stance and Voice in Written Academic Genres*. edited by K. Hyland, and C. S. Guinda, 101–115. London, UK: Palgrave Macmillan.

Burton, M., and C. Morgan. 2000. "Mathematicians Writing." *Journal for Research in Mathematics Education* 31, no. 4: 429–453. https://doi.org/10.2307/749652.

Chan, T. H. T. 2015. "A Corpus-Based Study of the Expression of Stance in Dissertation Acknowledgements." *Journal of English for Academic Purposes* 20: 176–191. https://doi.org/10.1016/j.jeap.2015.09.005.

Cheng, F. W., and L. Unsworth. 2016. "Stance-Taking as Negotiating Academic Conflict in Applied Linguistics Research Article Discussion Sections." *Journal of English for Academic Purposes* 24: 43–57. https://doi.org/10.1016/j.jeap.2016.09.001.

Conrad, S., and D. Biber. 2000. "Adverbial Marking of Stance in Speech and Writing." In *Evaluation in Text: Authorial Distance and the Construction of Discourse*, edited by S. Hunston and G. Thompson, 56–73. Oxford, UK: Oxford University Press.

Dang, T. N. Y. 2018. "The Nature of Vocabulary in Academic Speech of Hard and Soft-Sciences." *English for Specific Purposes* 51: 69–83. https://doi.org/10.1016/j.esp.2018.03.004.

Dreschler, G., M. Ontrust, and N. de Jong. 2019. "Digital Individual Support for Writing Skills Across the Disciplines." Paper Presented at the 2019 Conference for the European Association for the Teaching of Academic Writing. https://easychair.org/smart-program/EATAW2019/2019-07-03.html#talk:106283.

Egbert, J. 2015. "Sub-Register and Discipline Variation in Published Academic Writing: Investigating Statistical Interaction in Corpus Data." *International Journal of Corpus Linguistics* 20, no. 1: 1–29. https://doi.org/10.1075/ijcl.20.1.01egb.

Ehret, K., and M. Taboada. 2021. "The Interplay of Complexity and Subjectivity in Opinionated Discourse." *Discourse Studies* 23, no. 2: 141–165. https://doi.org/10.1177/1461445620966923.

Fitzmaurice, S. 2004. "Subjectivity, Intersubjectivity and the Historical Construction of Interlocutor Stance: From Stance Markers to Discourse Markers." *Discourse Studies* 6: 427–448. https://doi.org/10.1177/1461445604046585.

Fox, J., and S. Weisberg. 2018. "Visualizing Fit and Lack of Fit in Complex Regression Models With Predictor Effect Plots and Partial Residuals." *Journal of Statistical Software* 87, no. 9: 1–27. https://doi.org/10.18637/jss.v087.i09.

Fox, J., and S. Weisberg. 2019. *An R Companion to Applied Regression*, Third Edition, LA, US: Sage.

Gray, B. 2015. *Linguistic Variation in Research Articles: When Discipline Tells Only Part of the Story*. Amsterdam, NL: John Benjamins.

Gray, B., and D. Biber. 2012. "Current Conceptions of Stance." In *Stance and Voice in Written Academic Genres*, edited by K. Hyland and C. S. Guinda, 15–33. New York, US: Palgrave Macmillan.

Gray, B., and D. Biber. 2015. "Stance Markers." In *Corpus Pragmatics: A Handbook*, edited by K. Aijmer and C. Rühlemann, 219–248. Cambridge, UK: Cambridge University Press.

Gross, A. G., and P. Chesley. 2012. "Hedging, Stance and Voice in Medical Research Articles." In *Stance and Voice in Written Academic Genres*, edited by K. Hyland and C. S. Guinda, 85–100. New York, US: Palgrave Macmillan.

Halliday, M. A. K. 1988. "On the Language of Physical Science." In *Registers of Written English*, edited by M. Ghadessy, 162–172. London, UK: Pinter.

Hardy, J. A., and U. Römer. 2013. "Revealing Disciplinary Variation in Student Writing: A Multi-Dimensional Analysis of the Michigan Corpus of Upper-Level Student Papers (MICUSP)." *Corpora* 8, no. 2: 183–207. https://doi.org/10.3366/cor.2013.0040.

Harrell, Jr., and E. Frank. 2022. "*rms: Regression Modeling Strategies*." R Package Version 6.3-0. https://CRAN.R-project.org/package=rms.

Hothorn, T., P. Buehlmann, S. Dudoit, A. Molinaro, and M. Van Der Laan. 2006. "Survival Ensembles." *Biostatistics* 7, no. 3: 355–373. https://doi.org/10.1093/biostatistics/kxj011.

Huang, M., and Y. Chang. 2008. "Characteristics of Research Output in Social Sciences and Humanities: From a Research Evaluation Perspective." *Journal of the American Society for Information Science and Technology* 59, no. 11: 1819–1828. https://doi.org/10.1002/asi.20885.

Hunston, S. 1994. "Evaluation and Organization in a Sample of Written Academic discourse." In *Advances in Written Text Analysis*, edited by M. Coulthard, 191–218. London, UK: Routledge.

Hunston, S., and G. Thompson. 2000. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford, UK: Oxford University Press.

Hyland, K. 1999. "Disciplinary Discourses: Writer Stance in Research Articles." In *Writing: Texts, Processes and Practices*, edited by C. N. Candlin and K. Hyland, 99–121. London, UK: Routledge.

Hyland, K. 2004. "Disciplinary Interactions: Metadiscourse in L2 Postgraduate Writing." *Journal of Second Language Writing* 13, no. 2: 133–151. https://doi.org/10.1016/j.jslw.2004.02.001.

Hyland, K. 2005. "Stance and Engagement: A Model of Interaction in Academic Discourse." *Discourse Studies* 7, no. 2: 173–192. https://doi.org/10.1177/1461445605050365.

Hyland, K. 2012. "Undergraduate Understandings: Stance and Voice in Final Year Reports." In *Stance and Voice in Written Academic Genres*, edited by K. Hyland and C. S. Guinda, 134–150. New Yok, US: Palgrave Macmillan.

Hyland, K., and F. Jiang (Kevin). 2017. ""We Believe That … ": Changes in an Academic Stance Marker." *Australian Journal of Linguistics* 38, no. 2: 139–161. https://doi.org/10.1080/07268602.2018.1400498.

Hyland, K., and F. Jiang. 2016. "Change of Attitude?" *A Diachronic Study of Stance. Written Communication* 33, no. 3: 251–274. https://doi.org/10.1177/0741088316650399.

Hyland, K., and H. J. Zou. 2021. ""I Believe the Findings are Fascinating": Stance in Three-Minute Theses." *Journal of English for Academic Purposes* 50: 100973. https://doi.org/10.1016/j.jeap.2021.100973.

Hyland, K., and P. Tse. 2005a. "Hooking the Reader: A Corpus Study of Evaluative *That* in Abstracts." *English for Specific Purposes* 24, no. 2: 123–139. https://doi.org/10.1016/j.esp.2004.02.002.

Hyland, K., and P. Tse. 2005b. "Evaluative *that* Constructions: Signalling Stance in Research Abstracts." *Functions of Language* 12, no. 1: 39–63. https://doi.org/10.1075/fol.12.1.03hyl.

Kim, C., and P. Crosthwaite. 2019. "Disciplinary Differences in the Use of Evaluative *That*: Expression of Stance via *that*-Clauses in Business and Medicine." *Journal of English for Academic Purposes* 41: 100775. https://doi.org/10.1016/j.jeap.2019.100775.

Levshina, N. 2015. *How to Do Linguistics With R. Data Exploration and Statistical Analysis*. Amsterdam, NL: John Benjamins.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik Maechler. 2019. "cluster: Cluster Analysis Basics and Extensions." R Package Version 2.1.0. https://CRAN.R-project.org/package=cluster.

Martin, J. R., and P. R. White. 2005. *The Language of Evaluation: Appraisal in English*. New York, US: Palgrave Macmillan.

Napolitano, D. M., and A. Stent. 2009. "TechWriter: An Evolving System for Writing Assistance for Advanced Learners of English." *Calico Journal* 26, no. 3: 611–625.

Pendar, N., and E. Cotos. 2008. "Automatic Identification of Discourse Moves in Scientific Article Introductions." In Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications, edited by J. Tetreault, J. Burstein, and R. De Felice, 62–70. Association for Computational Linguistics. https://www.aclweb.org/anthology/W08-0908.pdf.

Qiu, X., and F. K. Jiang. 2021. "Stance and Engagement in 3MT Presentations: How Students Communicate Disciplinary Knowledge to a Wide Audience." *Journal of English for Academic Purposes* 51: 100976. https://doi.org/10.1016/j.jeap.2021.100976.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org.

Salager-Meyer, F., M. Á. A. Ariza, and M. L. Briceño. 2012. "The Voice of Scholarly Dispute in Medical Book Reviews, 1890–2010." In *Stance and Voice in Written Academic Genres*, edited by K. Hyland and C. S. Guinda, 232–248. New York, US: Palgrave Macmillan.

Storer, N. W. 1967. "The Hard Sciences and the Soft: Some Sociological Observations." *Bulletin of the Medical Library Association* 55, no. 1: 75–84.

Suzuki, R., Y. Terada, and H. Shimodaira. 2019. "pvclust: Hierarchical Clustering with p-Values via Multiscale Bootstrap Resampling." R Package Version 2.2-0. http://stat.sys.i.kyoto-u.ac.jp/prog/pvclust/.

Tagliamonte, S. A., and R. H. Baayen. 2012. "Models, Forests, and Trees of York English: *Was/Were* Variation as a Case Study for Statistical Practice." *Language Variation and Change* 24, no. 2: 135–178. https://doi.org/10.1017/S0954394512000129.

Thompson, D. K. 1993. "Arguing for Experimental "Facts" in Science: A Study of Research Article Results Sections in Biochemistry." *Written Communication* 10, no. 1: 106–128. https://doi.org/10.1177/0741088393010001004.

Thompson, P. 2012. "Achieving a Voice of Authority in PhD theses." In *Stance and Voice in Written Academic Genres*, edited by K. Hyland and C. S. Guinda, 119–133. New York, US: Palgrave Macmillan.

Thompson, P., and S. Hunston. 2019. *Interdisciplinary Research Discourse: Corpus Investigations Into Environment Journals*. London, UK: Routledge.

Tse, P. 2012. "Stance in Academic bios." In *Stance and Voice in Written Academic Genres*, edited by K. Hyland and C. S. Guinda, 69–84. New York, US: Palgrave Macmillan.

Veirano Pinto, M. 2014. "Dimensions of Variation in North American movies." In *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, edited by T. Berber Sardinha and M. Veirano Pinto, 109–147. Amsterdam, NL: John Benjamins.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics With S*. Cham, Switzerland: Springer.

Wang, F., and I. Pramoolsook. 2021. "Attitude in Abstracts: Stance Expression in Translation Practice Reports and Interpretation Practice Reports by Chinese Students." *Discourse and Interaction* 14, no. 1: 100–123. https://doi.org/10.5817/DI2021-1-100.

Winkworth, T., T. Tredgold, and J. Glynn. 1858. "Journal of the Society of Arts." *The Journal of the Society of Arts* 6, no. 310: 697–706.

Wu, B., and B. Paltridge. 2021. "Stance Expressions in Academic Writing: A Corpus-Based Comparison of Chinese Students' MA Dissertations and PhD Theses." *Lingua* 253: 103071. https://doi.org/10.1016/j.lingua.2021.103071.

Xie, J., J. Xie, and G. Bui. 2024. "A Diachronic Study of Authorial Stance in the Discussion of Chinese MA Theses and Published Research Articles." *Journal of English for Academic Purposes* 67: 101320. https://doi.org/10.1016/j.jeap.2023.101320.

Yang, W. 2016. "Evaluative Language and Interactive Discourse in Journal Article Highlights." *English for Specific Purposes* 42: 89–103. https://doi.org/10.1016/j.esp.2016.01.001.

Zou, H. J., and K. Hyland. 2022. "How the Medium Shapes the Message: Stance in Two Forms of Book Reviews." *Journal of Pragmatics* 193: 269–280. https://doi.org/10.1016/j.pragma.2022.03.023.

Appendix 1: List of Stance Features

CATEGORY	LINGUISTIC UNITS	SOURCE
Modal and semi-modal verbs		Biber (2006)
• Possibility/permission/ability	*Can*, *could*, *may*, *might*	
• Necessity/obligation	*Must*, *should*, *(had) better*, *have to*, *got to*, *ought to*	
• Prediction/volition	*Will*, *would*, *shall*, *be going to*	
Stance adverbs		
• Epistemic: certainty	*Actually*, *always*, *certainly*, *definitely*, *indeed*, *inevitably*, *in fact*, *never*, *of*	
• Epistemic: likelihood	*apparently*, *evidently*, *kind of*, *in most cases/instances*, *perhaps*, *possibly*, *predictably*, *probably*, *roughly*, *sort of*, *maybe*	
• Attitude	*Amazingly*, *astonishingly*, *conveniently*, *curiously*, *hopefully*, *even worse*, *fortunately*, *importantly*, *ironically*, *rightly*, *sadly*, *surprisingly*, *unfortunately*	
• Style	*According to*, *confidentially*, *frankly*, *generally*, *honestly*, *mainly*, *technically*, *truthfully*, *typically*, *reportedly*, *primarily*, *usually*	
Stance complement clauses controlled by verbs		
• Stance verb + *that*-clause, e.g., *the research showed that*		
◦ Epistemic verbs: certainty	*Conclude*, *demonstrate*, *determine*, *discover*, *find*, *know*, *learn*, *mean*, *notice*, *observe*, *prove*, *realize/realise*, *recognize/recognise*, *remember*, *see*, *show*, *understand*	
◦ Epistemic verbs: likelihood	*Assume*, *believe*, *doubt*, *gather*, *guess*, *hypothesize/hypothesise*, *imagine*, *predict*, *presuppose*, *presume*, *reckon*, *seem*, *speculate*, *suppose*, *suspect*, *think*	
◦ Attitude verbs	*Agree*, *anticipate*, *complain*, *concede*, *ensure*, *expect*, *fear*, *feel*, *forget*, *hope*, *mind*, *prefer*, *pretend*, *require*, *wish*, *worry*	
◦ Speech act and other communication verbs	*Announce*, *argue*, *assert*, *claim*, *contend*, *declare*, *emphasize/emphasise*, *explain*, *imply*, *insist*, *mention*, *promise*, *propose*, *recommend*, *remark*, *respond*, *say*, *state*, *suggest*, *tell*	
• Stance verb + *to*-clause, e.g., *we decided to analyse*		
◦ Probability (likelihood) verbs	*Appear*, *happen*, *seem*, *tend*	
◦ Cognition/perception verbs (likelihood)	*Assume*, *believe*, *consider*, *expect*, *find*, *forget*, *imagine*, *judge*, *know*, *learn*, *presume*, *pretend*, *remember*, *suppose*	
◦ Desire/intention/decision verbs	*Agree*, *choose*, *decide*, *hate*, *hesitate*, *hope*, *intend*, *like*, *love*, *mean*, *need*, *plan*, *prefer*, *prepare*, *refuse*, *want*, *wish*	
◦ Verbs of causation/modality/effort	*Allow*, *attempt*, *enable*, *encourage*, *fail*, *help*, *instruct*, *manage*, *oblige*, *order*, *permit*, *persuade*, *prompt*, *require*, *seek*, *try*	
◦ Speech act and other communication verbs	*Ask*, *claim*, *invite*, *promise*, *remind*, *request*, *be said*, *show*, *teach*, *tell*, *urge*, *warn*	
Stance complement clauses controlled by adjectives		
• Stance adjective + *that*-clause, e.g., *it is apparent that*		

(Continues)

CATEGORY	LINGUISTIC UNITS	SOURCE
◦ Epistemic adjectives: certainty	*Apparent*, *certain*, *clear*, *confident*, *convinced*, *correct*, *evident*, *false*, *impossible*, *inevitable*, *obvious*, *positive*, *right*, *sure*, *true*, *well-known*	
◦ Epistemic adjectives: likelihood	*Doubtful*, *likely*, *possible*, *probable*, *unlikely*	
◦ Attitude/emotion adjectives	*Afraid*, *amazed*, *aware*, *concerned*, *disappointed*, *encouraged*, *glad*, *happy*, *hopeful*, *pleased*, *shocked*, *surprised*, *worried*	
◦ Evaluation adjectives	*Amazing*, *appropriate*, *conceivable*, *crucial*, *essential*, *fortunate*, *imperative*, *inconceivable*, *incredible*, *interesting*, *lucky*, *necessary*, *nice*, *noteworthy*, *odd*, *ridiculous*, *strange*, *surprising*, *unacceptable*, *unfortunate*	
• Stance adjective + *to*-clause, e.g., *it is necessary to mention*		
◦ Epistemic adjectives: certainty	*Certain*, *guaranteed*, *sure*	
◦ Epistemic adjectives: likelihood	*Apt*, *due*, *liable*, *likely*, *prone*, *unlikely*	
◦ Attitude/emotion adjectives	*Afraid*, *ashamed*, *disappointed*, *embarrassed*, *glad*, *happy*, *pleased*, *proud*, *puzzled*, *relieved*, *sorry*, *surprised*, *worried*	
◦ Evaluation adjectives	*(In)appropriate*, *bad/worse*, *good/better/best*, *convenient*, *essential*, *important*, *interesting*, *necessary*, *nice*, *reasonable*, *silly*, *smart*, *stupid*, *surprising*, *useful*, *useless*, *unreasonable*, *wise*, *wrong*	
◦ Ability or willingness adjectives	*(Un)able*, *anxious*, *careful*, *determined*, *eager*, *eligible*, *hesitant*, *inclined*, *obliged*, *prepared*, *ready*, *reluctant*, *(un)willing*	
◦ Ease or difficulty adjectives	*Difficult*, *easier*, *easy*, *hard*, *(im)possible*, *tough*	
Stance complement clauses controlled by nouns		
• Stance noun + *that*-clause, e.g., *it's a fact that*		
◦ Epistemic nouns: certainty	*Assertion*, *conclusion*, *conviction*, *discovery*, *doubt*, *fact*, *knowledge*, *observation*, *principle*, *realization/realisation*, *result*, *statement*	
◦ Epistemic nouns: likelihood	*Assumption*, *belief*, *claim*, *contention*, *feeling*, *hypothesis*, *idea*, *implication*, *impression*, *notion*, *opinion*, *possibility*, *presumption*, *suggestion*	
◦ Attitude/perspective nouns	*Grounds*, *hope*, *reason*, *view*, *thought*	
◦ Communication (non-factual) nouns	*Comment*, *news*, *proposal*, *proposition*, *remark*, *report*, *requirement*	
• Stance noun + *to*-clause, e.g., *we had an intention to start with*	*Agreement*, *decision*, *desire*, *failure*, *inclination*, *intention*, *obligation*, *opportunity*, *plan*, *promise*, *proposal*, *reluctance*, *responsibility*, *right*, *tendency*, *threat*, *wish*, *willingness*	
Hedges	*Assume*, *broadly*, *certain amount/extent/level*, *claim*, *could*, *doubt*, *doubtful*, *essentially*, *estimate*, *fairly*, *feel*, *frequently*, *from our/this perspective*, *generally*, *guess*, *indicate*, *in general*, *in most cases/instances*, *in my/our opinion/view*, *largely*, *likely*, *mainly*, *may*, *maybe*, *might*, *mostly*, *often*, *on the whole*, *ought*, *perhaps*, *plausible*, *plausibly*, *possible*, *possibly*, *postulate*, *presumable*, *probable*, *probably*, *quite*, *rather*, *relatively*, *roughly*, *seems*, *should*, *sometimes*, *somewhat*, *suggest*, *suppose*, *suspect*, *tend to*, *to my knowledge*, *typical*, *typically*, *uncertain*, *uncertainly*, *unclear*, *unclearly*, *unlikely*, *usually*, *would*	Hyland (2005)

(Continues)

Category	Linguistic units	Source
	Attempt, *certain degree*, *can*, *especially*, *from my point of view*, *kindly*, *kind of*, *particularly*, *properly*, *simply*, *to some/certain extent*	Wang and Pramoolsook (2021)
Boosters	*Demonstrate*, *doubtless*, *establish*, *evident*, *evidently*, *in fact*, *incontestable*, *incontrovertible*, *incontrovertibly*, *indeed*, *indisputable*, *indisputably*, *know*, *must*, *never*, *no doubt*, *obvious*, *obviously*, *of course*, *prove*, *realize/realise*, *really*, *show*, *sure*, *surely*, *think*, *truly*, *true*, *undeniable*, *undeniably*, *undisputedly*, *undoubtedly*, *unexpected*, *without doubt*	Hyland (2005)
	Hold, *it is well known*, *point out*, *shall*, *will*	Wang and Pramoolsook (2021)
Attitude markers	*Correctly*, *curious*, *curiously*, *desirable*, *desirably*, *disappointed*, *disappointing*, *disappointingly*, *disagree*, *dramatic*, *dramatically*, *essential*, *essentially*, *even*, *expected*, *expectedly*, *fortunate*, *fortunately*, *hopeful*, *hopefully*, *important*, *importantly*, *inappropriate*, *inappropriately*, *interesting, interestingly*, *prefer*, *preferable*, *preferably*, *remarkable*, *remarkably*, *shock*, *shocking*, *shockingly*, *striking*, *strikingly*, *surprised*, *surprising*, *surprisingly*, *unbelievable*, *unbelievably*, *understandable*, *understandably*, *unexpected*, *unexpectedly*, *unfortunate*, *unfortunately*, *unusual*, *unusually*, *usual*	Hyland (2005)
	Crucial, *easily*, *effective*, *great*, *hope*, *practical*, *significant*, *significance*, *skillfully*	Wang and Pramoolsook (2021)
Self-mention	*I*, *me*, *mine*, *our*, *the author*, *the author's*, *the writer's*, *we*	Hyland (2005)
	He, *she*	Wang and Pramoolsook (2021)

Appendix 2: Normalised Frequency of Stance Features per 100,000 Words

Stance feature	Chemistry	Engineering	Maths	Physics	*Hard total*	Business	History	Linguistics	Political science	*Soft total*	Chi-squared	*p* value	
Modal verbs: possibility	468.1	555.9	430.3	697.7	*538.3*	737.5	333.01	588.07	562.7	*555.1*	0.03	0.866	n/s
Modal verbs: necessity	73.8	97.5	83.7	112.8	*92.05*	101.03	112.3	132.3	173.7	*130.06*	41.6	< 0.0001	***
Modal verbs: prediction	105.6	143.2	252.7	221.7	*181.08*	199.03	270.3	191.05	279.1	*235.01*	41.6	< 0.0001	***
Stance adverbs: certainty	49.04	52.7	93.2	100.9	*74.09*	82.8	151.4	136.8	112.8	*121.07*	77.2	< 0.0001	***
Stance adverbs: likelihood	23.7	20.1	27.5	27.4	*24.7*	26.7	76.7	73.1	47.1	*56.1*	87.8	< 0.0001	***

(Continues)

STANCE FEATURE	CHEMISTRY	ENGINEERING	MATHS	PHYSICS	*HARD TOTAL*	BUSINESS	HISTORY	LINGUISTICS	POLITICAL SCIENCE	*SOFT TOTAL*	CHI-SQUARED	*p* VALUE	
Stance adverbs: attitude	25.2	9.5	4.01	7.5	*11.5*	14.6	24.5	17.9	24.2	*20.3*	16.5	0.00005	***
Stance adverbs: style	97.5	122.1	37.6	61.4	*79.6*	103.05	84.7	117.9	108.3	*103.5*	18.6	0.00002	***
That-clauses controlled by adjs: certainty	4.04	10.05	13.5	15.4	*10.8*	5.05	20.5	9.9	11.3	*11.7*	0.07	0.7978	n/s
That-clauses controlled by adjs: likelihood	7.5	3.5	0.5	1.5	*3.2*	6.5	11.03	11.9	9.4	*9.7*	23.04	< 0.0001	***
That-clauses controlled by adjs: attitude	0.5	0	0	0	*0.1*	0.5	2.5	4.9	0.9	*2.2*	—	0.00008	***
That-clauses controlled by adjs: evaluation	4.5	1.01	1.00	4.5	*2.7*	3.03	4.01	1.9	2.4	*2.8*	0	1	n/s
To-clauses controlled by adjs: certainty	0	2.5	0	1.5	*1*	0	0	0	0.9	*0.2*	—	0.1001	n/s
To-clauses controlled by adjs: likelihood	14.1	3.5	0	5.5	*5.7*	79.8	18.05	19.4	131.1	*62.2*	356.2	< 0.0001	***
To-clauses controlled by adjs: attitude	1.5	0	0	0	*0.3*	0.5	1.00	2.4	0	*0.9*	—	0.2540	n/s
To-clauses controlled by adjs: evaluation	16.6	22.1	13.04	55.9	*26*	26.2	31.09	22.8	26.7	*26.2*	0.3	0.5471	n/s
To-clauses controlled by adjs: ability	38.9	20.6	15.5	23.4	*24.6*	37.8	41.6	33.3	35.1	*36.9*	16.2	0.00006	***
To-clauses controlled by adjs: difficulty	27.3	36.7	32.1	50.4	*36.6*	25.7	26.08	25.8	25.2	*25.7*	18.07	0.00002	***
That-clauses controlled by verbs: certainty	115.2	188.5	314.9	252.2	*218*	169.2	79.7	140.8	198.4	*147.1*	129.4	< 0.0001	***
That-clauses controlled by verbs: likelihood	14.6	40.7	126.8	51.9	*58.6*	30.8	35.6	36.8	41.08	*36.1*	48.7	< 0.0001	***
That-clauses controlled by verbs: attitude	9.1	15.5	17.05	17.4	*14.8*	22.2	18.5	16.4	28.2	*21.3*	7.4	0.0065	**
That-clauses controlled by verbs: speech act	57.1	47.2	149.9	60.4	*78.7*	171.7	158.4	152.7	184.1	*166.7*	227.3	<0.0001	***

(Continues)

STANCE FEATURE	CHEMISTRY	ENGINEERING	MATHS	PHYSICS	*HARD TOTAL*	BUSINESS	HISTORY	LINGUISTICS	POLITICAL SCIENCE	*SOFT TOTAL*	CHI-SQUARED	*p* VALUE	
To-clauses controlled by verbs: probability	19.7	17.5	25.08	23.9	*21.6*	47.9	43.1	70.6	64.3	*56.6*	114.6	< 0.0001	***
To-clauses controlled by verbs: cognition	0	1.01	0	8.4	*2.3*	2.02	0	0.5	0.4	*0.7*	6.2	0.0124	*
To-clauses controlled by verbs: desire	0	1.01	1	5.9	*2.01*	3.5	1	1.4	0.9	*1.7*	0.08	0.7710	n/s
To-clauses controlled by verbs: causation	0.5	1.5	1	7.9	*2.7*	8.08	4.01	2.9	4.4	*4.8*	3.5	0.0518	n/s
To-clauses controlled by verbs: speech act	0.5	0	0	0.5	*0.2*	0.5	0	1	0.4	*0.5*	—	0.6879	n/s
That-clauses controlled by nouns: certainty	16.1	27.6	57.1	26.4	*31.1*	21.7	26.08	43.7	21.7	*28.3*	1.8	0.1773	n/s
That-clauses controlled by nouns: likelihood	8.09	4.5	19.5	7.4	*9.9*	30.8	23.5	24.8	37.1	*29.1*	69.1	< 0.0001	***
That-clauses controlled by nouns: attitude	1.01	0	0	1	*0.5*	7.5	9.5	8.9	4.4	*7.6*	—	< 0.0001	***
That-clauses controlled by nouns: communication	1.01	2.5	10.03	4.9	*4.6*	3.5	2.01	5.9	8.4	*5*	0	0.9567	n/s
To-clauses controlled by nouns	5.5	4.5	5.5	5.9	*5.4*	35.3	39.6	37.8	41.08	*38.4*	188.2	< 0.0001	***
Hedges	624.8	826.9	921.3	983.4	*839.7*	1062.3	707.6	947.2	1066.6	*946.7*	25.2	< 0.0001	***
Boosters	5.1	5.2	10.6	7.04	*1.7*	15.1	11.2	15.6	11.3	*3.3*	27.3	< 0.0001	***
Attitude markers	117.2	57.8	53.6	95.4	*81*	111.1	103.3	90.5	209.8	*128.9*	74.1	< 0.0001	***
Self-mention	566.2	501.6	1991.1	1140.8	*1051.3*	995.1	320.9	704.9	1070.5	*773.3*	414.9	< 0.0001	***