

Article

From Corpus to Curriculum: Disciplinary Variation in Doctoral Abstracts and Its Implications for Writing Instruction

Ying Chai

Yingqi Huang

The Hong Kong Polytechnic University, China

Yueyue Huang

Fujian Normal University, China

Jing Deng*

The Hong Kong Polytechnic University, China

Received: 21 April 2026/Received in revised form: 16 September 2026/Accepted: 2 October 2026/Available online: 30 October 2026

Abstract

A recurring challenge in English for Academic Purposes (EAP) writing instruction is deciding how far a single, shared curriculum can serve doctoral writers from different disciplines, and where teaching must instead be calibrated to particular fields. Drawing on a corpus of 800 English doctoral abstracts submitted with Chinese doctoral dissertations in Biology, Chemistry, Physics, and Engineering, this exploratory study examined 20 lexical and syntactic complexity indices to identify where disciplines differ or show limited differentiation. The Entropy Weight Method (EWM) estimated each index's distributional salience, while multivariate and univariate analyses assessed disciplinary differentiation. Disciplinary effects were broader and generally larger among the lexical indices than among the syntactic indices. Chemistry had the highest means on several lexical sophistication and diversity indices, while Physics showed a narrower pattern centered on verbal and modifier variation. Syntactic differentiation was more limited: coordinate phrases per T-unit (CP/T) was the only syntactic index significant after Holm adjustment, and conclusions from the non-inferiority analyses depended on a post-specified effect-size bound. The findings provide a provisional rationale for combining shared instruction in abstract organization and information packaging with sustained, discipline-specific vocabulary work; they do not establish a universally sufficient syntactic common core.

Keywords

Disciplinary variation, linguistic complexity, Entropy Weight Method, English doctoral abstracts, Chinese doctoral writers

*Corresponding author. Email: jing.deng@cpce-polyu.edu.hk

1 Introduction

Doctoral EAP instructors face a practical problem when they design curriculum. Administrators and program designers, often under institutional pressure to run a single course for a linguistically and disciplinarily diverse doctoral cohort, look for a common core that can be taught across fields (Hyland, 2002). However, the common core is hard to define. When only the frequency of linguistic forms is considered, it is often possible to identify features that appear shared across disciplines; whether this apparent similarity reflects a shared communicative function, however, cannot be established without also examining what these forms are doing within each disciplinary context. The result raises a question less of either-or than of degree: not whether writers from different disciplines should be taught through a common syllabus or discipline-specific instruction, but where the boundary between them lies, which apparently shared forms may support a shared communicative function across disciplines, and at what point surface similarity gives way to functional divergence that calls for disciplinary calibration. This tension is particularly salient in doctoral abstracts: as condensed representations of disciplinary conventions, they require novice scholars to adapt their linguistic choices to discipline-specific communicative expectations while meeting the dual demands of clarity and economy (Biber & Gray, 2010; Hu & Bonsu, 2025; Nasseri & Thompson, 2021). In this sense, patterns of linguistic choice extend beyond stylistic embellishment; rather, they reflect the disciplinary norms and practices that guide communication (Hao et al., 2026; Hyland, 2015). Hard disciplines are characterized by knowledge construction that is linear, problem-oriented, and relatively standardized, with established procedures for producing and validating claims (Hyland, 2004, 2008, 2015; Storer, 1967). Within this, however, Biology, Chemistry, Physics, and Engineering differ in their objects of inquiry, technical naming demands, and preferred ways of organizing evidence. These differences may shape linguistic choice at the lexical level, through the naming of entities and processes specific to each discipline, and at the syntactic level, through the grammatical organization of technical information. Yet doctoral abstracts are also governed by shared genre pressures: their strict length constraints and dense, self-contained nature require writers to package information economically (Ansarifar et al., 2018), while their disciplinary readership presupposes familiarity with specialized communicative practices (Hyland, 2008). Linguistic complexity therefore offers a useful lens for examining how discipline-specific knowledge construction and shared abstract-genre constraints jointly shape communicative choices in doctoral writing (Kim et al., 2025). In effect, it can test the curriculum-design question raised above.

Given these two forces, we hold tentative expectations, rather than formal hypotheses, about where disciplinary differences are most likely to surface. Knowledge-construction accounts (Hyland, 2004, 2008) suggest that fields differ both in what they must name and in how they relate entities, materials, organisms, and processes to one another. At the lexical level, academic vocabulary varies across disciplines in frequency, meaning, and collocation (Hyland & Tse, 2007), so lexical measures may register disciplinary differences. However, the four fields examined here are all hard-science disciplines sharing a scientific register, and the abstract is a highly conventionalized part-genre; lexical differences may therefore be relatively modest. At the syntactic level, the same length constraints press writers toward compact packaging, and in academic writing such compression is achieved largely through phrasal rather than clausal elaboration (Biber & Gray, 2010; Ansarifar et al., 2018). Which phrasal and clausal resources a field favors may nonetheless differ, and earlier studies report syntactic differences that are selective, concentrated in particular measures (Khany & Kafshgar, 2016). We therefore expect disciplinary differences at both levels, but tentatively anticipate that syntactic measures will be the more sensitive, with greater sensitivity understood as larger and more consistent disciplinary effects across indices in the dimension-specific analyses. This expectation concerns disciplinary differentiation, not the weight an index carries in the composite complexity model. Because differences may be selective and need not appear across all disciplinary pairs (W. Zhou et al., 2023), we make no prediction about which pairs of disciplines will differ.

Previous research on disciplinary variation has often focused on linguistic features, such as lexical bundles (Durrant, 2017; Hyland, 2008) and metadiscourse nouns (Jiang & Hyland, 2017), with limited attention to whether multidimensional patterns of lexical and syntactic complexity retain measurable disciplinary variation within the shared constraints of the abstract genre. To address this gap, the present study measures the lexico-syntactic complexity of English abstracts written by Chinese doctoral students using an established inventory derived from L2 complexity research (e.g., Bulté & Housen, 2012; Lu, 2010), rather than a set of features selected or “identified” by any single theoretical framework. These indices provide common descriptive measures through which disciplinary differentiation and cross-disciplinary similarity can be assessed at the lexical and syntactic levels, with the expectations set out above treated as tentative rather than as hypotheses. The knowledge construction perspective provides the macro-level rationale for interpreting both lexical and syntactic patterning, while the Register-Functional approach is used more narrowly to explain how the observed syntactic patterns are structurally realized, especially through different clausal and phrasal resources (Biber et al., 2021). The lexical patterns, by contrast, are interpreted mainly in relation to discipline-specific naming demands and knowledge construction practices.

Methodologically, the present study combines correlation-based statistical screening, the Entropy Weight Method (EWM), overall and dimension-specific MANOVAs, and follow-up univariate ANOVAs. Correlation-based screening and EWM are used to retain and weight indices for the composite linguistic complexity model, whereas MANOVAs and ANOVAs are used to examine disciplinary differentiation. This design makes it possible to ask, for each index, not only whether disciplines differ and how much each index matters to complexity overall, but also how an observed difference might be realized in text, through illustrative extracts that suggest possible functions without establishing them. Therefore, the present study contributes to research on Chinese doctoral writers’ English academic writing in three ways: (1) it examines linguistic complexity in doctoral abstracts through a weighting-based lens, asking which indices carry the most distributional information in the corpus and how this salience relates to disciplinary differentiation rather than treating all indices as equally informative; (2) it offers functionally motivated interpretations, illustrated with concrete examples, of how specific forms may construct disciplinary meaning; and (3) it outlines corpus-informed implications for doctoral EAP writing instruction, offered as priorities for curriculum design rather than as tested recommendations.

In this study, disciplinary knowledge construction and shared genre constraints are assumed to shape lexical and syntactic choices, but EWM-based salience and discipline-based differentiation are treated as related rather than equivalent properties.

2 Literature Review

2.1 Disciplinary knowledge construction under genre and register constraints

Disciplinary knowledge construction refers to the field-specific ways in which communities generate, organize, validate and disseminate claims (Hyland, 2004, 2015), which are expected to leave measurable traces in linguistic choice. Recent research has begun to examine it through multidimensional patterns of linguistic complexity (Hao et al., 2026; Yi & He, 2026). Building on this perspective, recurrent patterns of lexical and syntactic complexity are treated here as potential linguistic realizations of how academic communities shape and communicate knowledge, rather than as direct measures of disciplinary epistemology (Hyland, 2004). Because linguistic complexity is itself a multidimensional construct rather than a single continuum from less to more complex (Bulté & Housen, 2012), disciplinary differentiation need not be distributed uniformly across its component measures (Casal et al., 2021; Ziaeeian & Golparvar, 2022): lexical research commonly distinguishes density, sophistication, and variation, whereas syntactic research has examined production-unit length, subordination, coordination, and phrase-level sophistication (Kim, 2021; Kyle & Crossley, 2018; Lu, 2010; Norris & Ortega, 2009).

The available evidence on lexical complexity points to multidimensional and selective disciplinary variation. Early comparisons found lexical density and diversity to vary independently rather than in parallel: density has emerged as the stronger discriminator in some studies (Khany & Kafshgar, 2016), while diversity has done so in others (Ayadi, 2024). At a larger scale, Zhu et al. (2024) found a general diachronic increase in lexical density across 16 disciplines, with hard-science abstracts denser than soft-science ones, while Pan and Yang (2025) showed that density, sophistication, and diversity all increased over time in Economics and Medicine yet remained discipline-sensitive throughout. Broadening the comparison across ten humanities and social-science disciplines, Hao et al. (2026) likewise found that density, sophistication, and diversity produced partially distinct disciplinary profiles.

A similarly selective pattern holds for syntactic complexity, though the location of disciplinary effects shifts with the measures and textual units examined. Khany and Kafshgar (2016) found that subordination ratio, but not average sentence length, differentiated their three disciplines, suggesting that effects concentrate in some structural measures rather than others. Register-Functional research has extended such comparisons by situating syntactic structures within rhetorical context: syntactic-rhetorical relationships have been shown to vary by discipline in research article introductions (Lu et al., 2021) and discussions (Ziaeian & Golparvar, 2022), while comparisons across six STEM disciplines found complexity varying by rhetorical move-step but stable between hard-pure and hard-applied fields (W. Zhou et al., 2023). Using mean dependency distance, Yi and He (2026) found disciplinary contrasts concentrated around Materials Engineering, with Biology instead overlapping Applied Linguistics and Sociology. Although not equivalent to the indices used here, these findings motivate asking whether disciplinary effects in the present study are broadly distributed or localized across the four syntactic dimensions.

Genre and register further condition the realization of disciplinary patterns (Biber & Conrad, 2009; Biber et al., 2021; Ziaeian & Golparvar, 2022). Findings obtained from full research articles or individual sections therefore cannot be transferred straightforwardly to doctoral abstracts. El-Dakhs (2018), for example, identified systematic differences in rhetorical organization and metadiscourse between PhD thesis abstracts and research article abstracts, indicating that the two are governed by distinct genre expectations. Within dissertation abstracts, Nasser and Thompson (2021) further showed that lexical density and diversity remain empirically distinguishable and respond differently to writers' language backgrounds despite the compressed textual format. As a highly compressed and conventionalized part-genre, the doctoral abstract therefore creates a shared communicative environment that may constrain the range of lexico-grammatical choices without necessarily eliminating variation (Paltridge & Starfield, 2020; Swales & Feak, 2009). From this perspective, both disciplinary differentiation and cross-disciplinary similarity are theoretically plausible outcomes.

2.2 Lexical and syntactic complexity in doctoral abstracts

As reviewed in Section 2.1, research on linguistic complexity has focused primarily on published research articles, whereas doctoral theses and their abstracts have received less systematic attention. Within disciplinary communities, doctoral theses serve as an important mechanism for the dissemination and ratification of knowledge (Koutsantoni, 2006), occupying a position between student coursework and published expert writing and exhibiting their own rhetorical and linguistic conventions (Xiao & Sun, 2020). Existing studies of doctoral texts have mainly examined rhetorical organization (El-Dakhs, 2018), value arguments (Hu & Bonsu, 2025), and authorial stance (Zhao, 2025). Research addressing linguistic complexity in postgraduate dissertations has remained comparatively limited and has generally focused on differences in academic level (Anjarwati et al., 2024) or writers' language backgrounds (Nasser and Thompson, 2021; Wang & Lowie, 2021), rather than disciplinary variation. More specifically, studies of English doctoral abstracts written by Chinese writers have generally examined either lexical bundles

(Bao & Liu, 2023; Lu & Deng, 2019) or syntactic complexity (Wang & Lowie, 2021), often through comparisons between writer groups rather than among closely related scientific disciplines.

Few studies have therefore examined lexical and syntactic complexity together across a group of closely related scientific disciplines. It remains unclear whether disciplinary differentiation in English doctoral abstracts written by Chinese writers extends broadly across lexical and syntactic dimensions or is localized in a limited number of measures under the shared constraints of the abstract genre. The present study addresses this gap by applying a common inventory of text-level indices to doctoral abstracts from Biology, Chemistry, Physics, and Engineering. It distinguishes the corpus-wide distributional salience of the retained indices from their discipline-based effects, interprets both lexical and syntactic patterns in relation to disciplinary knowledge construction, and uses the Register-Functional (RF) perspective more specifically to explain how the observed syntactic patterns are structurally realized.

2.3 Research questions

The relevant question is therefore not simply whether one discipline is more complex than another, but which dimensions vary across disciplines and under what textual conditions. The study is exploratory: the tentative expectations set out in the Introduction guide interpretation but are not tested as formal hypotheses. In light of the research gaps discussed above, the present study aims to investigate disciplinary variations in the linguistic complexity of English doctoral abstracts written by Chinese writers across four hard disciplines. The specific research questions addressed are as follows:

- RQ1: What disciplinary variations can be observed in the lexical and syntactic complexity of English doctoral abstracts written by Chinese writers across four hard disciplines?
- RQ2: Which lexico-grammatical features carry the largest weights in the composite linguistic complexity score under the Entropy Weight Method?
- RQ3: To what extent do the feature weights derived from EWM correspond to the disciplinary effects estimated by ANOVA across the four hard disciplines?

3 Method

3.1 Corpus compilation

Data for this study were retrieved from the China Doctoral Dissertations Full-text Database (CDFD), hosted by the China National Knowledge Infrastructure (CNKI), covering the period from 2005 to 2020. The starting year was set at 2005 because data for certain disciplines prior to this year were relatively limited, which could potentially compromise the reliability of the statistical analyses. To avoid the period of widespread access to generative AI (GenAI) tools, an upper limit of 2020 was adopted; widespread public use of such tools accelerated after the release of ChatGPT in November 2022 (Picazo-Sanchez & Ortiz-Martin, 2024). This reduces but does not eliminate the possibility of assisted writing. English texts by Chinese scholars are often shaped by supervisors, language teachers, editors, or translators (Burrough-Boenisch, 2003; Luo & Hyland, 2016), and machine translation can raise lexical sophistication and syntactic complexity in L2 texts (Chon et al., 2021). Because such assistance could not be verified for individual abstracts and may differ across disciplines, the indices are interpreted as characteristics of the abstracts as submitted, not of the writers' unaided competence. Given that the texts collected in this study are all English abstracts submitted by Chinese doctoral writers, the compiled corpus was named the Chinese Doctoral Abstracts (CDA) corpus.

Based on the disciplinary classification within the CDA, four disciplines were selected: Biology, Chemistry, Physics, and Engineering. These four fields are commonly associated with the broader STEM (Science, Technology, Engineering, and Mathematics) domain, though this grouping does not reflect

a deliberate attempt to sample one discipline from each STEM subcategory: subject categories within CNKI's doctoral dissertation database do not map cleanly onto the four STEM letters (e.g., "Technology" and "Engineering" are not consistently distinguished as separate categories). The four disciplines are therefore better characterized, and are referred to throughout this study, as hard disciplines in Hyland's (2008) sense, which share commonalities in knowledge construction characterized by a linear, problem-oriented approach and a reliance on standardized, efficient forms of argumentation. These shared characteristics, however, provide a relatively controlled foundation for examining disciplinary variation. At the same time, hard disciplines are not entirely homogeneous, as previous research suggests that certain pairs of these fields pattern together along different lexical and syntactic dimensions, with the specific configurations varying by feature (Yang et al., 2026; Yi & He, 2026). Such partially overlapping affinities suggest that these four disciplines, while comparable as hard disciplines, may pattern differently across lexical and syntactic dimensions, a configuration that is well suited to the multidimensional weighting approach adopted in the present study. Moreover, these four disciplines have been examined in previous research on linguistic complexity (Lu et al., 2021; Pérez-Guerra & Smirnova, 2023; Yi & He, 2026; Ziaieian & Golparvar, 2022), thereby providing an empirical foundation for the current study.

Before sampling, records were screened for the availability of an English abstract, textual completeness, disciplinary relevance, and obvious formatting problems. Disciplinary relevance and textual completeness were independently checked by two doctoral students. Records receiving discrepant judgments were conservatively excluded rather than treated as eligible. Where exclusions left fewer than 200 eligible abstracts in a discipline, additional records were retrieved from CNKI using the same disciplinary search strategy and eligibility criteria, screened in the same manner, and added until the target size was reached; this supplementation was not stratified by publication year. No fixed word-length cutoff was imposed. The final corpus ranged from 150 to 1,626 words (Table 1), and only one abstract (chemistry_84, 1,626 words) exceeded 1,500 words; it was retained because it was complete and relevant to the target discipline. From each initial eligible pool, abstracts were randomly selected using a random number generator (Urbaniak & Plous, 2013) to minimize selection bias; supplementary records, where needed, followed the same eligibility criteria but were not randomly sampled or stratified by year.

Table 1
Description of the Corpus

Subject	No. of Abstracts	Length Range (words)	Mean Length (words)	Type	Token	STTR
Biology	200	173–1,101	599.23	12,601	119,845	40.97
Chemistry	200	276–1,626	609.08	12,820	121,816	42.41
Engineering	200	159–922	575.5	9,981	115,100	39.78
Physics	200	150–1,018	534.97	10,100	106,993	41
Total	800			28,669	463,754	

Note: TYPE refers to the number of unique word forms within each disciplinary subcorpus. The total TYPE value was calculated from the full 800-abstract corpus and is therefore not the sum of the four disciplinary TYPE counts. STTR (standardized type-token ratio) was calculated on the basis of 1,000 words, following the method adapted from Kwok et al. (2025). Word counts treat alphanumeric strings separated by whitespace or punctuation as individual tokens.

3.2 Linguistic complexity measures

During the data collection and preprocessing stage, all texts were saved in plain text (.txt) format. Some crawled records lacked corresponding English abstracts, contained omitted sentences, or did not match

the target discipline because of technical limitations in automatic data acquisition and database access. These checks are described in Section 3.1. To reduce length-related bias, the study excluded raw-count and strongly length-sensitive measures, and retained ratio-based or intra-unit indices that normalize complexity by lexical, clausal, or sentential units rather than by absolute text size (Lei, 2017; Lu, 2012).

Traditionally, linguistic complexity has often been operationalized using syntax-focused tools such as the L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010). While these tools are widely used to extract syntactic features, they do not compute lexical indices. Because the present study examines doctoral abstracts across both lexical and syntactic dimensions, NeoSCA (version 0.1.4; Tan, 2024), a reimplementation of Lu's (2010) L2SCA and Lu's (2012) Lexical Complexity Analyzer that favors Stanza-based parsing in releases after 0.1.0, was selected as the primary analysis tool. All 800 abstracts were processed as separate plain-text files under the same English-language configuration, without discipline-specific rules or manual correction. The indices are therefore treated as automated estimates of the established L2SCA and LCA measures, which have been applied to academic and doctoral-abstract writing (e.g., Wang & Lowie, 2021). Equipped with built-in dictionaries, part-of-speech taggers, and syntactic parsers, NeoSCA can extract multidimensional linguistic features on a unified computing platform, thereby reducing inconsistencies in tokenization standards and data alignment that may arise when separate tools are used. Potential redundancy among fine-grained indices was addressed through the selection procedure described in Section 3.3.

3.3 Selection of complexity indices

Index selection and weighting proceeded in two EWM stages. First, length-sensitive indices were removed, highly collinear indices were dropped (Spearman $r > .80$), and a preliminary EWM score served as the target for a stepwise regression and VIF check that reduced the candidate set to 20 indices. Second, EWM was recalculated on these 20 indices to weight the LC score. The regression statistics and formulas are given in Appendix A.

To improve the accuracy and interpretability of the results, this study adopted a strategy integrating theoretical classification with data-assisted approaches for index selection. The procedure involved three stages. First, absolute counts and length-sensitive indices (e.g., types, tokens, NDW, TTR, W, S, VP, C, T, DC, CT, CP, CN) were excluded because of their sensitivity to text length. Second, because several indices violated the assumption of normality, Spearman correlation analysis was conducted in SPSS to remove highly collinear variables (threshold set at $r > .80$). Third, a regression-based redundancy diagnostic was conducted. The preliminary EWM-derived composite score was used as an auxiliary screening target to identify indices that retained non-redundant explanatory contribution after collinearity control. Because this score was constructed from the same candidate indices, the regression step was not interpreted as external predictive validation of the complexity construct. Instead, stepwise regression and iterative VIF checks were used as a statistical device for reducing redundancy among overlapping complexity measures. Any index yielding a VIF value above 10.0 was removed, and the model was re-evaluated until all retained predictors were statistically significant ($p < .05$) and below the VIF threshold. Ultimately, 20 indices were retained. The lexical dimension encompasses lexical density (LD), lexical sophistication (LS1, LS2, VS1), and lexical variation (MSTTR, Uber, NDW-50, VV1, VV2, SVV1, AdvV, ModV) (Bui, 2021; Bulté & Housen, 2012; Zhang & Lu, 2024). Meanwhile, the syntactic dimension covers length of production unit (MLC), amount of subordination (C/T, CT/T, DC/C), amount of coordination (CP/T, T/S), and degree of phrase sophistication (CN/C, VP/T) (Dong et al., 2023; Lu, 2010). Because the retained indices may contribute unequally to disciplinary differentiation, an additional weighting procedure was applied.

3.4 The entropy weight method

As introduced in the Introduction, the Entropy Weight Method (EWM) was adopted to describe the distributional weight of each linguistic index within this corpus, based on its data dispersion. Rather than relying on subjective expert judgments, EWM assigns higher weights to indices whose normalized values are distributed more unevenly across texts (lower entropy), as these indices carry more distributional information in this corpus (Wang et al., 2015); it has also been applied in linguistic complexity research to evaluate multidimensional complexity indices across text types (Huang et al., 2025). Here, entropy is the Shannon (1948) information entropy of an index's normalized distribution across the 800 texts, used only as a measure of how unevenly that index is distributed; it is unrelated to thermodynamic entropy. The weight therefore describes the distribution of an index in this corpus; it does not indicate that the index is more important for writing quality, disciplinary differentiation, or teaching, and the LC score is a descriptive composite that was not validated against quality ratings. Because the weights depend on min–max normalization, indices with skewed distributions or extreme values may receive larger weights. Because the preliminary EWM score was used during screening, the final index set is path-dependent on this procedure; discipline labels, ANOVA results, and effect sizes were not used at any screening stage. An EWM-independent, VIF-only screening was therefore conducted as a sensitivity analysis (Appendix A). It retained 19 of the original 20 indices, replacing modifier variation (ModV) with adjective variation (AdjV); after EWM was recalculated, the alternative composite gave the same disciplinary ordering and a nearly identical disciplinary effect ($\eta^2 = .028$ vs. $.027$). In the primary procedure, EWM was used at two analytically distinct stages. After Spearman-based reduction, a preliminary EWM score was used only as an auxiliary target for redundancy screening. Because this score was a weighted linear combination of the same candidate indices, the regression model was not treated as independent criterion validation; its very high fit ($R^2 = .998$) was expected by design (Appendix A). A non-EWM composite computed as the unweighted mean of the normalized candidate indices correlated strongly with the preliminary EWM score (Pearson $r = .946$; Spearman $\rho = .936$; both $p < .001$) and produced the same discipline-level ordering (Chemistry > Physics > Biology > Engineering) at the candidate stage. For the final 20 indices, the EWM and unweighted composites likewise correlated strongly (Pearson $r = .945$; Spearman $\rho = .936$; both $p < .001$) and both showed an overall disciplinary effect, but the ordering of Biology and Engineering differed (Chemistry > Physics > Biology > Engineering under EWM; Chemistry > Physics > Engineering > Biology unweighted), and the effect was smaller under EWM ($\eta^2 = .027$) than under equal weighting ($\eta^2 = .066$; Appendix I). Chemistry's first and Physics' second position and the strong correlation are therefore stable under both weighting schemes, whereas the position of Biology and Engineering, the size of the composite disciplinary effect, individual weights, ranks, and EWM-based salience categories depend on the weighting procedure. The index-level results addressing RQ1 do not use EWM weights, although the VIF-only sensitivity analysis indicates that results involving ModV are dependent on which of two closely related indices is retained. In the second EWM stage, weights were recalculated only for the final 20-index subset to construct the LC score.

$$X_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

Standardization and entropy-based weighting followed the procedure detailed in Appendix A. Multiplying each text's standardized index values by their corresponding EWM weights and summing across indices yields each text's linguistic complexity (LC) score:

$$LC_i = \sum_{j=1}^n w_j x'_{ij}$$

Based on the final retained indices, the resulting LC model combines all 20 standardized indices as a weighted linear sum, with weights ranging from 0.025 (MSTTR) to 0.092 (CP/T); the specific weight assigned to each index is reported in Table 2.

Table 2
Calculations, Values, Weights, and Ranks of Linguistic Complexity Indices in CDA

Sub-category / Index	Calculation	Weight	Rank
Lexical density			
LD	N_{lex}/N	0.028	18
Lexical sophistication			
LS1	N_{slex}/N_{lex}	0.028	17
LS2	V_s/V	0.039	12
VS1	V_{sverb}/N_{verb}	0.037	13
Lexical variation			
MSTTR	Mean of TTRs for successive 50-word segments	0.025	20
Uber	$(\log N)^2/(\log N - \log V)$	0.045	10
NDW-50	Number of different words in the first 50 words	0.027	19
VV1	V_{verb}/N_{verb}	0.032	15
VV2	V_{verb}/N_{lex}	0.033	14
SVV1	T_{verb}^2/N_{verb}	0.047	8
AdvV	V_{adv}/N_{lex}	0.086	2
ModV	$(V_{adj} + V_{adv})/N_{lex}$	0.046	9
Length of production unit			
MLC	Number of words / Number of clauses	0.081	4
Amount of subordination			
C/T	Number of clauses / Number of T-units	0.030	16
CT/T	Number of complex T-units / Number of T-units	0.082	3
DC/C	Number of dependent clauses / Number of clauses	0.062	7
Amount of coordination			
CP/T	Number of coordinate phrases / Number of T-units	0.092	1
T/S	Number of T-units / Number of sentences	0.040	11
Degree of phrase sophistication			
CN/C	Number of complex nominals / Number of clauses	0.077	5
VP/T	Number of verb phrases / Number of T-units	0.063	6

Note: LD = lexical density; LS1 = lexical sophistication-I; LS2 = lexical sophistication-II; VS1 = verb sophistication-I; MSTTR = mean segmental type-token ratio; Uber = Uber index; NDW-50 = number of different words (50 words); VV1 = verb variation-I; VV2 = verb variation-II; SVV1 = squared verb variation-I; AdvV = adverb variation; ModV = modifier variation; MLC = mean length of clause; C/T = clauses per T-unit; CT/T = complex T-units per T-unit; DC/C = dependent clauses per clause; CP/T = coordinate phrases per T-unit; T/S = T-units per sentence; CN/C = complex nominals per clause; VP/T = verb phrases per T-unit. N = word tokens; V = word types; lex = lexical words; s = sophisticated. EWM weights are rounded to three decimal places, whereas ranks were calculated from the unrounded weights; consequently, indices with the same displayed weight may have different ranks.

3.5 Statistical and integrative analysis

To address RQ3, the final EWM weights reported in Table 2 were compared with the univariate η^2 values from the ANOVAs, reported in Appendices C and D. The EWM weights represent the contribution of each index to the composite LC model, a contribution that is sensitive to how the index is distributed, whereas η^2 indicates the extent to which each index differentiates the four disciplines. This comparison served to separate salience in the entropy model from differentiation across disciplines.

Each index was classified using two descriptive reference lines. Because EWM weights have no established convention for effect size, high salience was defined by the upper-quartile boundary of the 20 final weights (0.066), calculated through inclusive linear interpolation. For differentiation across disciplines, $\eta^2 = .060$ was adopted as the conventional benchmark for a medium effect (Cohen, 1988). These lines served as interpretive guides rather than formal cutoffs for inference, and they produced four categories: concordance (high EWM weight and high η^2); discordance dominated by the ANOVA (high η^2 but lower EWM weight); discordance dominated by EWM (high EWM weight but lower η^2); and cases below both lines, with low salience and low effect. These categories were subsequently interpreted within a two-level explanatory framework. Disciplinary knowledge construction served as the superordinate lens for both lexical and syntactic results. Because lexical complexity is multidimensional and the lexical findings were distributed across density, sophistication, and diversity, lexical patterns were interpreted directly in relation to disciplines' knowledge-construction demands. Syntactic patterns were interpreted through the more specific Register-Functional approach, which accounts for how such demands may be realized in clausal and phrasal structures.

Homogeneity of variance was assessed using median-centered Levene tests. Variances were unequal for LS1, LS2, MLC, CP/T, and CN/C, so conventional ANOVAs were supplemented with Welch ANOVAs (Delacre et al., 2019); post hoc comparisons used Tukey HSD when homogeneity was supported and Tamhane's T2 otherwise. Holm adjustment (Holm, 1979) was applied separately across the family of 20 conventional tests and the family of 20 Welch tests, and pairwise contrasts were interpreted only when the corresponding omnibus effect remained significant after adjustment. Partial eta squared (η^2), with 95% confidence intervals, was interpreted alongside p values. Residual Q-Q plots and discipline-specific boxplots were inspected; noticeable right-tail departures and isolated extreme observations were found for several syntactic ratio measures, particularly MLC, CP/T, T/S, CN/C, and VP/T. Holm-adjusted Kruskal-Wallis tests were therefore added as a rank-based robustness analysis. These reproduced the main dimensional pattern: all lexical indices except SVV1 showed disciplinary differences, whereas CP/T was the only syntactic index that remained significant. VV2 was significant in the rank-based analysis (Holm $p = .026$) but borderline in the Welch analysis (Holm $p = .052$) and is treated as procedure-sensitive (Appendix G). For the MANOVAs, Box's M was significant; because the four discipline groups were equal in size, Pillai's trace was used as the primary multivariate statistic. Formal multivariate-normality and Mahalanobis-distance screening was not conducted, so the MANOVA is interpreted together with the Welch and rank-based results. Finally, because a nonsignificant ANOVA does not establish similarity, omnibus non-inferiority tests were conducted for all eight syntactic indices following Campbell and Lakens (2021) and Larsson and Hancock (2026). The tests evaluated $H_0: \eta^2 \geq \Delta$ against $H_1: \eta^2 < \Delta$, with Holm adjustment across the eight indices. The primary sensitivity bound, $\Delta = .02$, was post-specified and lies between conventional small (.01) and medium (.06) reference values; alternative bounds of .01 and .03 were also examined. These analyses assess whether the overall disciplinary effect falls below a stated bound, not whether every pair of disciplines or every construction is equivalent (Appendix H).

3.6 Selection of illustrative extracts

Following a quantitatively informed illustration procedure adapted from Ädel (2014), the extracts were selected purposively after the statistical analysis. Selection focused on abstracts located toward the upper

end of the relevant index distributions, because such cases make the textual realization of a quantitatively identified feature more visible. Chemistry_139 ranked within the upper 10% of the Chemistry corpus on LS1, LS2, MSTTR, and Uber (Uber 1st, LS1 4th, LS2 18th, MSTTR 17th–18th of 200); physics_312 ranked within the top six on VV2 (1st), AdvV (6th), and ModV (6th); and biology_199 ranked 10th on CP/T. The index values were calculated for the complete abstracts, whereas the quoted passages represent only local illustrations. The extracts are therefore not treated as typical or representative texts and do not independently establish discipline-wide tendencies. Their rhetorical functions were not systematically coded across the wider corpus; the interpretations are presented as plausible text-level explanations rather than as independently established disciplinary tendencies.

4 Results

The overall MANOVA indicated a disciplinary effect on the 20 retained complexity indices. Because Box's M was significant (Box's M = 1229.855, $F = 1.868$, $p < .001$), Pillai's trace was reported as the primary multivariate statistic. The result was Pillai's trace = .458, $F(60, 2337) = 7.017$, $p < .001$, $\eta^2 = .153$ (Appendix B). Separate MANOVAs for the lexical and syntactic dimensions are also reported in Appendix B, followed by the corresponding univariate analyses in Sections 4.1 and 4.2. A single-case deletion check excluding the only abstract above 1,500 words (chemistry_84) left the significance status of all 20 indices unchanged, both before and after Holm adjustment. The absolute change in η^2 was below .002 for every index (largest: Uber, .0018), the CP/T effect remained .080, and the ordering of disciplinary means was unchanged for 19 of the 20 indices. The exception was the Chemistry–Engineering order on T/S, for which the disciplinary effect was nonsignificant. The substantive conclusions are therefore not attributable to this single long abstract.

4.1 Disciplinary variations in lexical complexity

At the lexical level, disciplinary variation was broad but uneven. As shown in Appendix C, 11 of the 12 lexical indices reached statistical significance, with SVV1 as the only nonsignificant index ($\eta^2 = .002$). After Holm adjustment, 11 lexical indices remained significant under conventional ANOVA and 10 under Welch ANOVA; VV2 was borderline in the Welch–Holm analysis ($p = .052$) and is therefore interpreted cautiously (Appendix G). The largest effects were observed for LS2 ($\eta^2 = .146$), LS1 ($\eta^2 = .128$), MSTTR ($\eta^2 = .113$), and Uber ($\eta^2 = .095$), indicating that disciplinary differences were concentrated mainly in lexical sophistication and diversity. Chemistry had higher values on lexical density (LD), lexical sophistication (LS1, LS2, and VS1), and several lexical variation indices (MSTTR, Uber, NDW-50, and VV1). Physics was more prominent on a narrower set of lexical variation indices, particularly VV2, AdvV, and ModV (Appendix E and Figure 1).

The extracts below illustrate how these lexical patterns can be realized in text. Example 1 (Chemistry) illustrates the results for LS1, LS2, MSTTR, and Uber, and Example 2 (Physics) those for VV2, AdvV, and ModV. Each was selected purposively from the upper end of the relevant index distributions (Section 3.6); neither is treated as a typical or representative text.

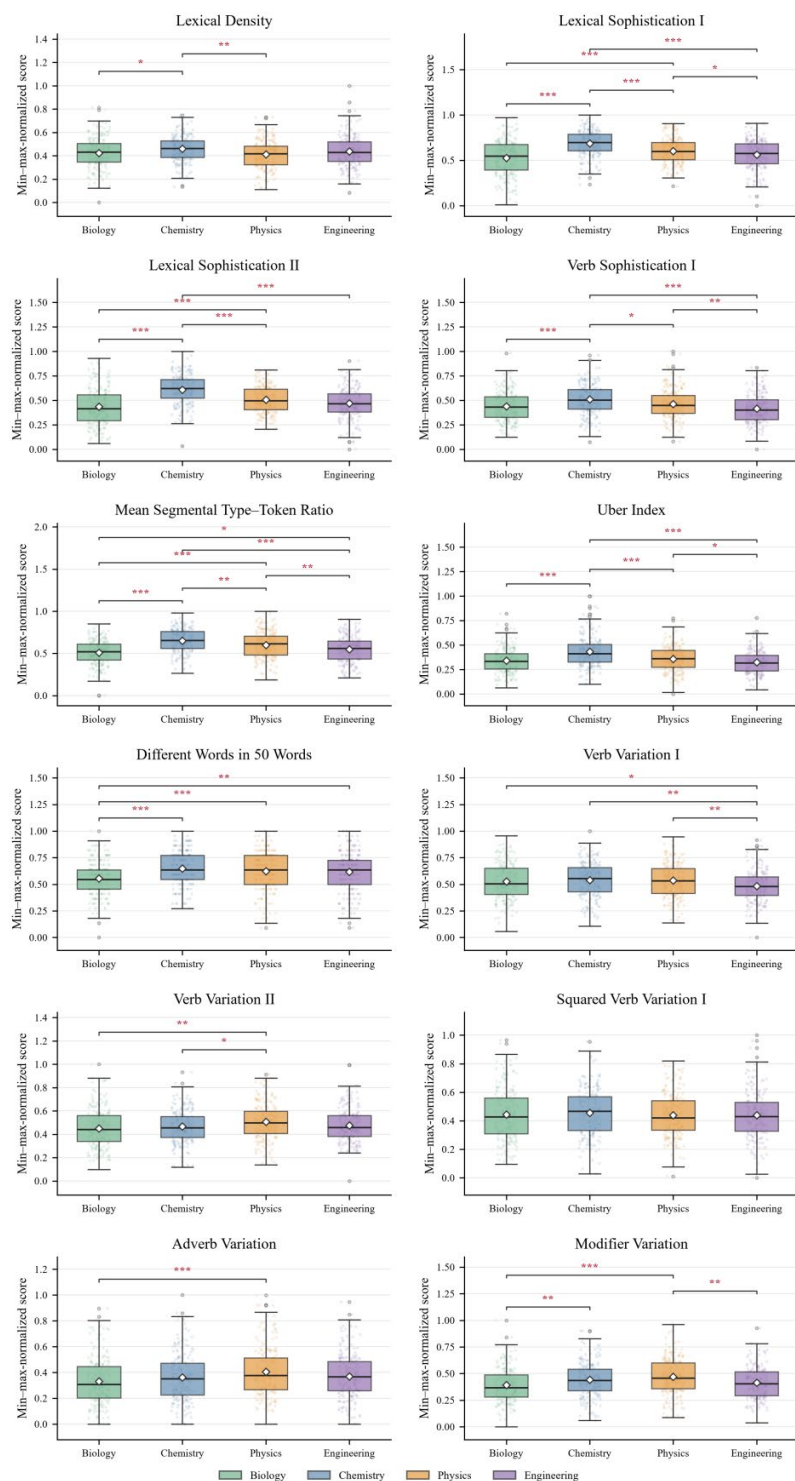
Example 1:

These compounds include thirty terpenoids, five polyketides, four long-chain fats, seven sterides, one glucoside, and one cyclic dipeptide. Among these compounds, compounds 1-31 were new compounds. Compounds 5 and 6 possessed a rare 5/6/4/5 fused tetracyclic ring skeleton and may be formed from compound 37 via an intramolecular [2+2] cycloaddition reaction. (chemistry_139.txt).

Example 2:

After going over its calculation, we find that one special Brownian motor named the Triangula would drift anomalously under some initial temperatures if the mass of particles in the two baths are not equal to each other, which is also supported by the stochastic simulation. By considering the physical realization in reality, we particularly discuss the motion of Brownian motors in one single bath with temperature gradient (physics_312.txt).

Figure 1

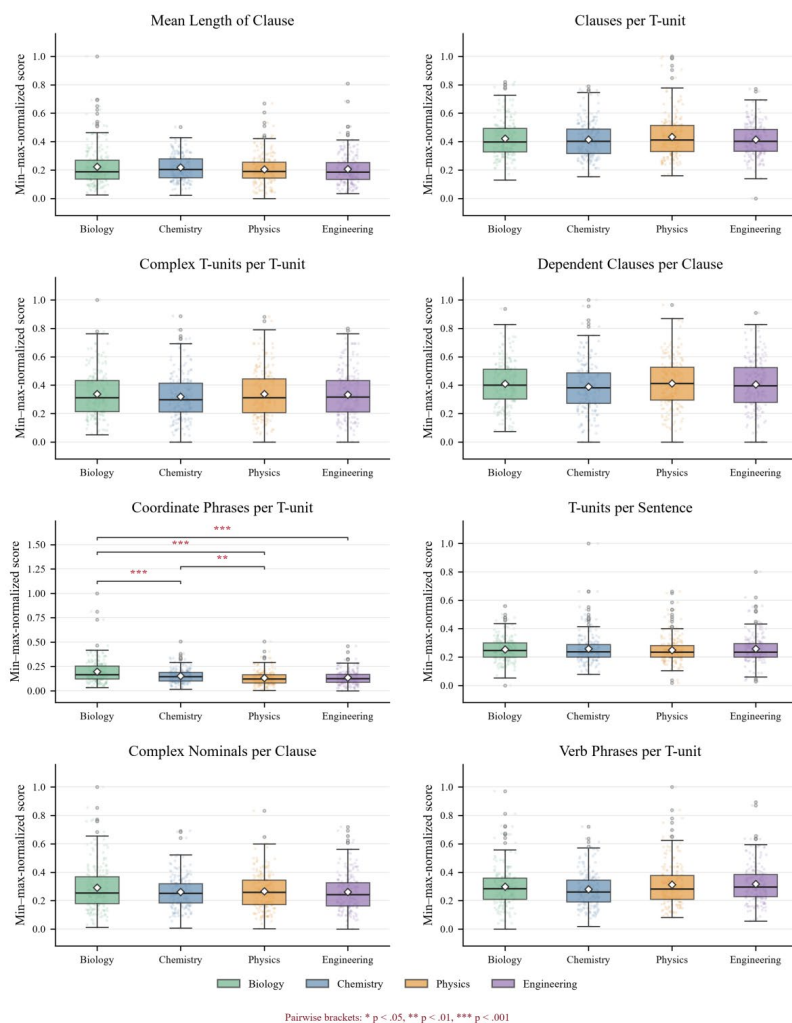
Detailed Lexical Complexity Comparisons

Note: Boxplots show min–max-normalized index scores. Pairwise contrasts were obtained using Tukey HSD when the homogeneity-of-variance assumption was supported and Tamhane’s T2 otherwise. Pairwise brackets are shown for indices whose conventional omnibus ANOVA remained significant after Holm adjustment. VV2 was significant in the conventional and rank-based analyses but borderline in the Welch–Holm analysis ($p = .052$) and is therefore interpreted cautiously.

4.2 Disciplinary variations in syntactic complexity

Figure 2

Detailed Syntactic Complexity Comparisons



Note: Boxplots show min–max-normalized index scores. Pairwise contrasts are displayed only for CP/T, the only syntactic index whose omnibus disciplinary effect remained significant after Holm adjustment. The contrasts were obtained using Tamhane’s T2 because the homogeneity-of-variance assumption was not met for CP/T. * $p < .05$, ** $p < .01$, *** $p < .001$.

Compared with lexical complexity, syntactic complexity showed fewer significant disciplinary differences (Appendix D and Figure 2). Among the eight syntactic indices, only CP/T remained significant after Holm adjustment; VP/T was significant before adjustment ($p = .024$) but not after it (Holm $p = .193$; Welch–Holm $p = .113$). CP/T, an index of coordination, showed the largest syntactic effect ($\eta^2 = .080$). Post hoc comparisons indicated that Biology was higher than Chemistry, Engineering, and Physics on CP/T; Chemistry was also higher than Physics ($p < .01$), although this secondary contrast

was smaller than Biology's contrasts with the other disciplines. The unadjusted Engineering–Chemistry contrast on VP/T (Appendix F) is not interpreted. Under the post-specified $\Delta = .02$ bound, omnibus non-inferiority tests supported effects below the bound for MLC, C/T, CT/T, DC/C, and T/S; CN/C and VP/T were inconclusive, and CP/T did not fall below the bound (Appendix H). The conclusions changed under alternative bounds, so these results are interpreted as sensitivity evidence rather than proof of syntactic equivalence.

Example 3 (Biology) illustrates the CP/T result; it was likewise selected from the upper end of the CP/T distribution (Section 3.6) and is not treated as a typical Biology abstract.

Example 3:

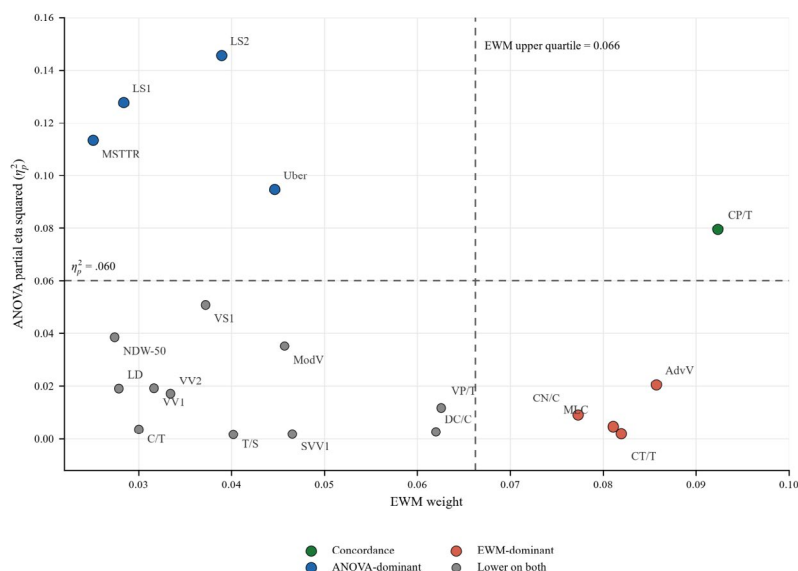
Ecosystems also perform fundamental life support services, which include the purification of air and water, detoxification and decomposition of wastes, regulation of climate, regeneration of soil fertility, and production and maintenance of biodiversity, mitigation of floods, droughts, the erosive forces of wind and rain (biology_199.txt).

4.3 Integrating EWM weights and disciplinary effects

Applying the criteria defined in Section 3.5, Figure 3 classifies the 20 retained indices into four integrative categories. Concordance was limited to CP/T, which combined the highest EWM weight (0.092, rank 1) with the strongest syntactic disciplinary effect reported in Section 4.2. Coordination therefore emerged as both highly salient in the entropy-based LC model and clearly discipline-sensitive.

Figure 3

EWM Weights and Disciplinary Effect Sizes Across the 20 Retained Indices



Note: The dashed vertical and horizontal lines indicate the upper-quartile EWM-weight threshold (.066) and the descriptive medium-effect reference value for partial eta squared ($\eta^2 = .060$), respectively. Both lines are descriptive guides rather than inferential cutoffs.

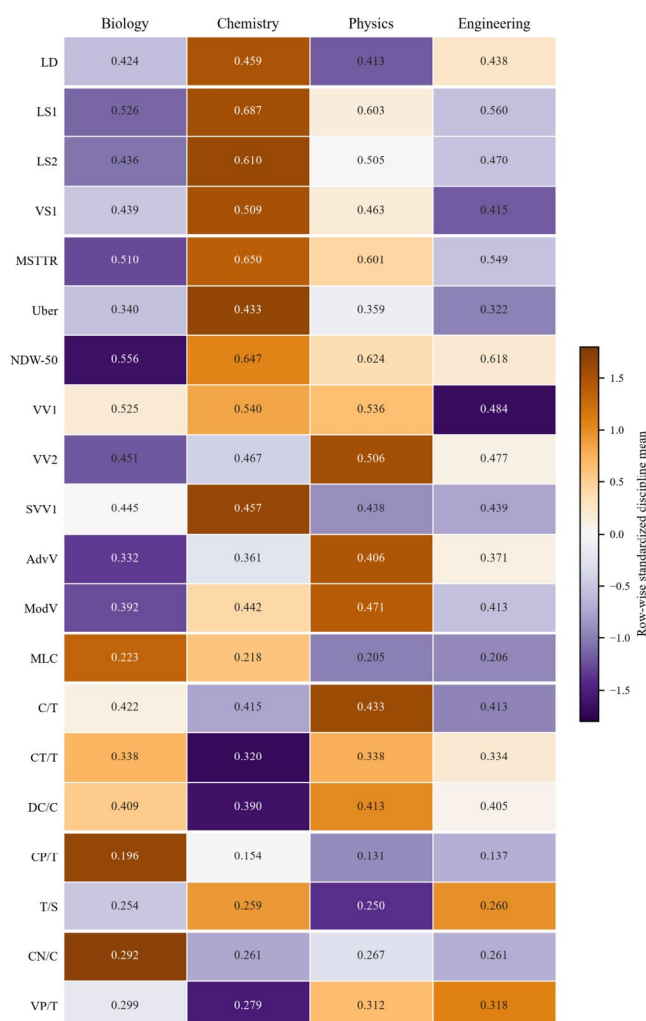
ANOVA-dominant discordance was observed for LS2, LS1, MSTTR, and Uber. As reported in Section 4.1, these lexical sophistication and diversity indices showed the strongest disciplinary effects, but their EWM ranks were only moderate or low (12th, 17th, 20th, and 10th, respectively). They are thus strong discipline markers without being the main contributors to the entropy-based composite model.

EWM-dominant discordance showed the reverse pattern. AdvV, MLC, CT/T, and CN/C ranked among the highest EWM-weighted indices, yet their disciplinary effects were small or non-significant, as shown in Sections 4.1 and 4.2. These indices contribute to overall LC variability, but that variability is not strongly organized by discipline.

The remaining 11 indices (LD, VS1, NDW-50, VV1, VV2, SVV1, ModV, C/T, DC/C, T/S, and VP/T) fell into the low-salience/low-effect category in Figure 3. They are retained in Figures 3 and 4 for completeness, but they do not define an additional integrative pattern. It should be noted that this category reflects effect sizes falling below the descriptive medium-effect threshold rather than an absence of statistical significance: six of these eleven indices (LD, VS1, NDW-50, VV1, VV2, and ModV) remained significant in the conventional univariate ANOVAs after Holm adjustment (Appendices C and G), whereas SVV1, C/T, DC/C, T/S, and VP/T did not.

As a complementary check on the weighting, the final 20 indices were also combined as an unweighted mean (Appendix I). The two composites correlated strongly (Pearson $r = .945$; Spearman $\rho = .936$) and both showed an overall disciplinary effect, but Biology and Engineering exchanged positions (EWM: Chemistry > Physics > Biology > Engineering; unweighted: Chemistry > Physics > Engineering > Biology), and the effect was smaller for the EWM composite ($\eta_p^2 = .027$) than for the unweighted composite ($\eta_p^2 = .066$). The Biology–Engineering order, the size of the composite effect, the weight ranks, and the salience categories should therefore be read as specific to the EWM procedure; the index-level results that address RQ1 do not depend on it.

Figure 4
Discipline-Level Mean Profiles of the 20 Retained Indices



Note: Cell entries are discipline-level means of the min–max-normalized index scores. Shading is based on row-wise standardized means and therefore indicates the relative disciplinary position within each index; color intensity should not be compared across different indices.

Figure 4 complements Figure 3 by showing the discipline-level mean profiles. Chemistry had the highest means on the clearest lexical discriminators, namely LS1, LS2, MSTTR, and Uber, with smaller supporting differences on VS1 and NDW-50. Physics showed a narrower profile centered on VV2, AdvV, and ModV, although VV2 was procedure-sensitive. Biology was distinguished most clearly by CP/T; its higher means on MLC and CT/T are not interpreted as disciplinary markers because neither index reached significance. Engineering did not form a distinctive profile: its highest VP/T mean did not survive Holm adjustment ($\eta p^2 = .012$) and is not interpreted.

5 Discussion

The central descriptive pattern is that disciplinary effects were broader and generally larger among the lexical indices than among the syntactic indices. LS1 and LS2 (lexical sophistication) and MSTTR and Uber (lexical diversity) differentiated the disciplines despite their moderate or low EWM weights. Chemistry had the highest means on these four indices, with LD and VS1 pointing in the same direction. One possible reading is that Chemistry abstracts must distinguish among closely related molecular entities, reaction mechanisms, and catalytic pathways, creating a naming demand consistent with Hyland's (2004, 2008) account of discipline-specific knowledge construction and with evidence that academic vocabulary varies across disciplines (Hyland & Tse, 2007). This reading remains tentative: LS1 and LS2 index rarity relative to general-English frequency lists, so specialist terms are likely to be counted as sophisticated irrespective of writing quality. Physics showed a narrower profile centered on VV2, AdvV, and ModV. Appendix E shows that AdvV differed only between Physics and Biology, VV2 between Physics and both Biology and Chemistry, and ModV between Physics and both Biology and Engineering; the VV2 result was borderline under Welch ANOVA after Holm adjustment ($p = .052$; Appendix G). This profile may relate to how conditions and manner are expressed in physical descriptions (Example 2), but that explanation was not tested directly. Engineering did not show a distinctive profile: it had the highest mean on no lexical index, and its highest VP/T mean did not survive Holm adjustment. Biology's higher CP/T may reflect a localized preference for enumeration (Example 3). Because all texts came from one genre, the smaller syntactic effects cannot be attributed causally to abstract-genre pressure; at most, they are compatible with the shared compactness requirements of doctoral abstracts.

Example 1 (Section 4.1), an extract from the upper end of the Chemistry distribution, illustrates how rare technical terms can realize the higher LS1 and LS2 observed for Chemistry. Specialized terms such as *terpenoids*, *polyketides*, *glucoside*, *cyclic dipeptide*, *tetracyclic ring skeleton*, and *intramolecular [2+2] cycloaddition reaction* are used to identify compound classes, structural configurations, and reaction mechanisms. A separate pattern of lexical variation is also illustrated in this example. Rather than repeatedly using a general superordinate term such as *natural products*, the first sentence distinguishes six compound categories. The concentration of distinct technical labels and structural descriptors within a short passage qualitatively illustrates the lexical variation quantitatively reflected in the higher MSTTR and Uber values for Chemistry abstracts. This finding complements previous research conducted at the syntactic level, which has shown that Chemistry and other hard-science disciplines tend to rely on complex nominals and phrasal structures to package specialized propositional content (Shen et al., 2023; Ziaeian & Golparvar, 2022). The present data are compatible with the view that Chemistry abstracts also package information through lexical sophistication and variation; however, CN/C did not differ across disciplines, so we do not claim that Chemistry shows more grammatical compression than the other fields.

The pattern illustrated in Example 2 adds nuance to previous findings that hard-science writing tends to favor phrasal and nominal resources for dense information packaging rather than extensive clausal elaboration (Biber & Gray, 2010; Biber et al., 2021; Staples et al., 2016). While previous research has highlighted noun-phrase compression as a defining feature of scientific discourse, the present findings suggest that disciplinary specificity may also emerge through lexical choices surrounding processes and conditions. In Example 2, higher VV2, AdvV, and ModV values indicate a greater diversification of verbs, adverbs, and modifiers that specify how physical processes occur, under what conditions they operate, and how theoretical predictions are evaluated. Thus, the present study does not challenge previous descriptions of scientific writing; rather, it extends this account by showing that within a generally compressed scientific register, physics doctoral abstracts may allocate precision to lexical resources associated with processes and conditions.

Despite the limited differentiation detected across most syntactic indices, CP/T constituted a notable exception, with Biology showing a significantly higher value than the other disciplines. Example 3 illustrates how coordinate structures organize multiple units of biological information within limited textual space. Although the finite relative clause “which include...” introduces the category of life-support services, the main informational burden is carried by a sequence of coordinated noun phrases and internally coordinated items, such as “the purification of air and water” and “detoxification and decomposition of wastes.” This pattern is consistent with Hyland’s (2008) characterization of Biology as a relatively discursive and descriptive discipline concerned with “naming and coding.” In this example, coordination allows multiple biological processes and ecological functions to be named and grouped within a single T-unit. The higher CP/T value therefore suggests that Biology may rely more strongly on coordinate phrases as a compact information-packaging resource. Because this extract was selected from the upper end of the CP/T distribution, it shows how coordination can be realized, not that it is typical of Biology abstracts. This finding adds nuance to W. Zhou et al.’s (2023) observation that syntactic complexity remained relatively stable across hard-pure and hard-applied science RA introductions. By comparing individual disciplines and focusing specifically on coordinate phrases in English doctoral abstracts written by Chinese writers, the present study identifies a finer-grained disciplinary difference that may be obscured by broader disciplinary groupings. Nevertheless, this CP/T difference represents a localized preference rather than broad syntactic divergence across the four disciplines.

The limited detected differentiation is better read not as an absence of syntactic patterning but as possible containment within one side of the Register-Functional elaboration–compression continuum (Biber & Gray, 2010; Biber et al., 2021). Little differentiation was detected in clausal elaboration: C/T, CT/T, DC/C, T/S, and MLC did not differ significantly across disciplines (all $\eta^2 \leq .005$), and the non-inferiority tests supported effects below the post-specified $\Delta = .02$ bound for each (Holm $p = .004$ –.027). No syntactic index passed at $\Delta = .01$, whereas seven of the eight passed at $\Delta = .03$ (Appendix H). These results are bound-dependent and do not imply identical use of individual constructions. CN/C ($p = .066$; non-inferiority Holm $p = .147$) and VP/T ($\eta^2 = .012$; Holm $p = .213$) remained inconclusive at $\Delta = .02$ and are not treated as showing similarity. This pattern may reflect the shared communicative demands of doctoral abstracts, which require substantial information to be packaged within limited textual space. Coordinate phrases can organize parallel elements compactly within T-units, but disciplinary variation at the syntactic level was concentrated in coordination (CP/T) rather than distributed across the broader set of measured resources. The genre-based interpretation remains tentative because no cross-genre comparison was conducted.

6 Implications

The asymmetry documented above offers a preliminary answer to the boundary question raised in the Introduction. The question is not whether shared or discipline-focused instruction should prevail, but where the boundary between them might be drawn. The findings indicate that disciplinary differentiation

in these doctoral abstracts was more extensive in lexis than in syntax. Pedagogically, this suggests that disciplinary calibration may be most warranted in lexical instruction, whereas syntax may provide a provisional shared starting point. Because the corpus comprises English-language doctoral abstracts retrieved from dissertations deposited in China, these implications speak most directly to doctoral EAP provision concerned with abstract writing, where novice scholars must learn to represent their research concisely and appropriately for specialist readers. The pedagogical recommendations below draw on effect sizes, corpus examples, communicative function, and pedagogical relevance rather than on EWM rank alone.

At the lexical level, a single general academic word list is unlikely to meet the different naming demands documented across the four disciplines. Chemistry's higher values for lexical sophistication and diversity on LS1, LS2, MSTTR, and Uber can be interpreted in relation to the need to distinguish molecular entities, structures, and reaction pathways precisely. This interpretation is consistent with Hyland and Tse's (2007) finding that academic words vary across disciplines in frequency, meaning, and collocation. LS1 and LS2 index rarity relative to general-English frequency lists and are not treated as measures of writing quality; specialist terms may be infrequent in general English because they perform necessary naming functions. The more selective profile found in Physics, centered on VV2, AdvV, and ModV, instead directs attention to how processes, conditions, manner, and modification are expressed. For learners, the goal should not be to increase lexical complexity scores or reproduce the profile of another discipline, but to develop the ability to select terms, collocations, and modifiers that express the intended disciplinary meaning accurately. Instructors can support this development by compiling small collections of expert abstracts and designing concordance activities in which students identify recurrent expressions, compare their functions across texts, and revise vague or inappropriate wording in their own abstracts. General resources such as the Academic Word List (Coxhead, 2000) can provide an initial foundation, but curriculum designers should supplement them with sustained vocabulary work based on texts from the learners' disciplines. The principle of converting corpus evidence into learner-facing resources has also been discussed in other instructional settings (Wong & Fong, 2025); that work concerns primary-school ESL learners and is cited here only for this general principle.

The corpus describes doctoral abstracts as submitted and is not an expert benchmark, so a feature that differs across disciplines is not thereby a deficiency or a teaching target. Teachers can treat a corpus finding as a candidate target when it also serves an identifiable communicative function, recurs in expert abstracts from the learners' own field (to be checked against a small reference collection), and is visibly problematic in learners' drafts, for example through vague naming or unclear enumeration. Higher complexity scores are not targets in themselves. Chemistry's higher LS1 and LS2 may largely reflect specialist terms that are infrequent in general English, so the aim is accurate naming of compounds and reactions, not more low-frequency words.

To illustrate, consider a Chemistry-focused activity on naming compounds and processes. First, the target is selected not because it receives high complexity scores but because it combines a notable disciplinary effect (here, LS1 and LS2; $\eta_p^2 = .128$ and $.146$), recurrence in the learners' disciplinary texts, an identifiable communicative function, and likely difficulty for learners. Second, in a concordance activity, students examine lines from a small collection of expert abstracts and identify what recurrent terms and noun phrases do in context. Third, in guided discussion, students compare more and less precise alternatives and consider when specialized terminology is required. Fourth, students revise a passage from their own abstract and justify their lexical choices. Progress is judged by lexical precision, collocational appropriateness, clarity, and suitability for disciplinary and genre purposes rather than by increases in LS, CP/T, or composite scores. A parallel sequence for Biology writers could compare coordinated noun phrases with clausal alternatives.

At the syntactic level, the limited differentiation detected across most measures, when interpreted alongside previous Register Functional research, is compatible with a shared genre pressure toward

compact information packaging (Ansarifar et al., 2018; Biber & Gray, 2010; Biber et al., 2011). This pattern provides a provisional rationale for shared instruction in resources such as nominalization, noun modification, complex nominals, and coordination, but it does not establish syntactic equivalence or a universally sufficient common core. The rationale rests mainly on the genre argument and on the five indices whose omnibus effects fell below the post-specified $\Delta = .02$ bound; CN/C, the index closest to complex nominals, was inconclusive, so complex nominals are a candidate shared target on pedagogical and genre grounds rather than on equivalence evidence. Two qualifications should guide this instruction. First, measures such as MLC and CN/C combine several structural features, so the absence of significant differences indicates limited detected differentiation rather than identical use of individual constructions. Second, Biology's higher CP/T demonstrates that writers in particular disciplines may favor different resources within the shared pressure for compression. Coordination can support compact enumeration in Biology, while examples from Chemistry may illustrate how technical information is packaged through specialized naming and nominal structures. Learners should therefore develop the ability to decide when information is most effectively expressed through coordination, nominal packaging, or clausal explanation. Teachers can facilitate this development by asking students to compare alternative formulations, evaluate their clarity and disciplinary function, and revise their own sentences rather than treating any single structure as a model to reproduce.

For curriculum designers, these findings support two parallel and connected strands rather than a sequence in which general instruction precedes disciplinary instruction. One strand can introduce all doctoral writers to the purposes of the abstract genre and to a repertoire of resources for compact and clear information packaging. The other can provide sustained lexical work and disciplinary examples from the beginning of the course, supported where possible by writing groups, subject specialists, and supervisor-guided revision. In practice, the shared strand can be taught to mixed-discipline cohorts by language teachers, while the disciplinary strand can be built from small reference collections of expert abstracts from learners' fields and supported by supervisors or subject specialists. The corpus does not indicate how time should be divided between the strands; that is a matter for local needs analysis. Diagnostic activities can connect the two strands by asking learners to compare their drafts with relevant corpus examples and justify their lexical and syntactic choices. Assessment should focus on lexical precision, collocational appropriateness, clarity, and suitability for disciplinary and genre purposes rather than on complexity values alone, for example by comparing learners' drafts and revised passages with reference abstracts against a short analytic rubric covering these criteria. Because the present study did not examine writers' instructional histories or test any teaching intervention, these recommendations should be treated as corpus-informed priorities for curriculum design rather than evidence that existing instruction produced the observed patterns. The findings therefore do not settle the common core question but draw a provisional and adjustable boundary between what doctoral writers across these four disciplines might productively learn together and what may require calibration to particular disciplinary contexts.

7 Conclusion

This study examined disciplinary variation in 20 lexical and syntactic complexity indices across English abstracts submitted with Chinese doctoral dissertations in four hard disciplines. EWM estimated the indices' distributional salience, while MANOVA and ANOVA assessed disciplinary differentiation; the Register-Functional approach was used to interpret the syntactic patterns. The largest disciplinary effects occurred in LS1 and LS2 (lexical sophistication) and MSTTR and Uber (lexical diversity), with Chemistry showing the highest values on these measures. Physics displayed a narrower pattern involving verb, adverb, and modifier variation (VV2, AdvV, and ModV). Syntactic differentiation was more limited: coordinate phrases per T-unit (CP/T) showed the only syntactic effect significant after Holm adjustment ($\eta^2 = .080$), distinguishing Biology from the other three disciplines. Under the post-specified $\Delta = .02$

bound, five other syntactic indices showed evidence of effects below the bound, while CN/C and VP/T were inconclusive; these conclusions changed at alternative bounds and do not establish equivalence. The comparison with EWM indicated that distributional salience and disciplinary differentiation were related but distinct. CP/T was the only index combining a high EWM weight with a medium-sized disciplinary effect, whereas several highly weighted indices showed weak or nonsignificant disciplinary differences. Academic writing complexity is therefore treated as multidimensional and register-sensitive rather than as a single, uniformly desirable property (Yasuda, 2024).

The findings are bounded to English abstracts submitted with Chinese doctoral dissertations, four hard disciplines, the selected lexical and syntactic indices, and the 2005–2020 period. Publication year was not used for stratification or retained as an observation-level covariate, and supplementary records added during sampling were not stratified by year, so possible diachronic change within the period was not isolated. The corpus provides a pre-ChatGPT baseline; it cannot establish current AI-assisted writing practices and cannot exclude earlier translation, editing, or other digital assistance, whose use for individual abstracts could not be verified (Burrough-Boenisch, 2003; Luo & Hyland, 2016; Chon et al., 2021). The indices are automated estimates: no discipline-stratified manual validation of tagging or parsing was conducted, and technical terms, abbreviations, chemical expressions, numerical identifiers, and dense noun phrases may produce discipline-dependent parsing errors, particularly for CP/T and CN/C. LS1 and LS2 reflect rarity relative to general-English lists, not writing quality. EWM was used in both screening and weighting, so the primary index set remains path-dependent; however, an EWM-independent VIF-only sensitivity analysis retained 19 of the 20 indices and produced the same composite ordering and a nearly identical disciplinary effect. Results involving ModV remain dependent on which of two closely related indices was retained. The relative position of Biology and Engineering, the size of the composite disciplinary effect, and individual EWM ranks also depended on the weighting rule. Non-significant ANOVA results do not establish equivalence: the non-inferiority analyses used a post-specified bound, their conclusions changed across $\Delta = .01, .02$, and $.03$, and the conventional ANOVA F statistics were used for all eight tests, including the heteroscedastic MLC, CN/C, and CP/T indices. Formal multivariate-normality and Mahalanobis-distance screening was not conducted. The illustrative extracts were selected purposively from the upper end of index distributions, and their rhetorical functions were not coded across the corpus. The study did not examine rhetorical organization, metadiscourse, other dissertation sections, or writers' instructional histories. Future research could validate the index reduction in an independent sample, conduct a discipline-stratified manual audit or use a second parser or dependency-based measures, and compare corpora compiled before and after the widespread adoption of generative AI.

Appendices

Appendix A.

Entropy weight method calculations

Since the selected linguistic indices possess different units and magnitudes, the original data matrix must be standardized into dimensionless values. Given an evaluation system with m texts and n indices, let x_{ij} represent the original value of the j^{th} index for the i^{th} text. All indices were direction-aligned as positive indicators (where higher values indicate greater complexity).

$$X_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

To prevent mathematical errors (i.e., taking the logarithm of zero in the subsequent step), a small constant of 0.0001 is added to all normalized values x'_{ij} .

Calculation of entropy

Next, the relative proportion y_{ij} of the i^{th} text for the j^{th} index is computed to determine the probability distribution:

$$y_{ij} = \frac{x'_{ij}}{\sum_{i=1}^m x'_{ij}}$$

Based on this proportion, the information entropy e_j of the j^{th} index is calculated:

$$e_j = -\frac{1}{\ln(m)} \sum_{i=1}^m y_{ij} \ln(y_{ij})$$

Where:

e_j = The entropy value for index j .

y_{ij} = The normalized proportion of sample i for index j .

x'_{ij} = The normalized value of sample i under index j .

m = The total number of samples.

$\ln(m)$ = A normalization factor to ensure that entropy values fall within the range $[0,1]$.

Determination of objective weights and final score

Finally, the objective weight w_j for each index is determined by its effective value $1-e_j$. The weight is calculated by normalizing the effective value against the sum of all indices:

$$\text{Effective value} = 1 - e_j$$

$$w_j = \frac{1 - e_j}{\sum_{j=1}^n (1 - e_j)}$$

The resulting weight w_j reflects the relative information content (degree of dispersion) of each index within this corpus; it does not indicate importance for writing quality.

Regression-based redundancy check. The final stepwise/VIF model retained 20 indices and reproduced almost all variance in the preliminary EWM score ($R^2 = .998$, adjusted $R^2 = .998$), with all VIF values below the exclusion threshold (range = 1.13–8.10). Because this score was a weighted linear combination of the same candidate indices, the high fit was expected by design and is interpreted only as evidence that the retained set preserved the information in the candidate pool after collinearity control, not as evidence of external predictive validity.

How screening may influence the composite. EWM was used at two stages, so the final index set is not fully independent of it. At the screening stage, the preliminary EWM score served as the target for stepwise regression. Because this score was a weighted combination of the candidate indices, the procedure could favor indices that contributed more uniquely to it and could affect which member of a highly correlated group was retained. This path dependence may shape the final index set and, in turn, the recalculated weights and composite scores. It cannot mechanically build in disciplinary differences, because discipline labels, ANOVA results, and effect sizes were not used at any screening stage.

Sensitivity to the screening route. Starting from the same 28 candidate indices that remained after theoretical, length-based, and Spearman-correlation screening, an EWM-independent procedure repeatedly removed the index with the highest VIF until all VIFs were below the same operational

threshold of 10 used in the primary analysis. This follows the iterative logic described by Zuur et al. (2010), while recognizing that 10 is a conventional rule rather than a universal cutoff (O'Brien, 2007). The alternative procedure also retained 20 indices, 19 shared with the primary set (Jaccard similarity = .905); adjective variation (AdjV) replaced modifier variation (ModV). After EWM was recalculated on the alternative set, its composite correlated almost perfectly with the original composite (Pearson $r = .9993$; Spearman $\rho = .9990$), produced the same disciplinary ordering (Chemistry > Physics > Biology > Engineering), and showed a nearly identical disciplinary effect, $F(3, 796) = 7.563$, $p < .001$, $\eta^2 = .028$, compared with $F(3, 796) = 7.382$, $p < .001$, $\eta^2 = .027$ for the original composite. The route affected which of two closely related indices was retained, but not the composite-level ordering or effect size. Index-level analyses in Section 4 use the primary 20-index set.

Appendix B. MANOVA summary for overall and dimension-specific analyses

Analysis	Dependent variables	Box's M	Box F	Box p	Pillai's Trace	F	df	p	η^2
Overall retained indices	20	1229.855	1.868	< .001	.458	7.017	60, 2337	< .001	.153
Lexical complexity	12	532.647	2.217	< .001	.348	8.593	36, 2361	< .001	.116
Syntactic complexity	8	354.645	3.226	< .001	.156	5.441	24, 2373	< .001	.052

Note: Box's M tests the equality of covariance matrices. Pillai's Trace is reported as the primary multivariate statistic because Box's M was significant and the four discipline groups had equal sample sizes ($n = 200$ per discipline). p values reported as .000 in SPSS are presented as < .001.

Appendix C. Descriptive statistics and one-way ANOVA results for lexical complexity indices across disciplines.

Index	Biology M (SD)	Chemistry M (SD)	Physics M (SD)	Engineering M (SD)	F(3, 796)	p	η^2
LD	0.424 (0.133)	0.459 (0.109)	0.413 (0.122)	0.438 (0.135)	5.158	.002	.019
LS1	0.526 (0.192)	0.687 (0.137)	0.603 (0.139)	0.560 (0.156)	38.839	< .001	.128
LS2	0.436 (0.190)	0.610 (0.152)	0.505 (0.135)	0.470 (0.151)	45.237	< .001	.146
VS1	0.439 (0.149)	0.509 (0.156)	0.463 (0.150)	0.415 (0.149)	14.243	< .001	.051
MSTTR	0.510 (0.147)	0.650 (0.135)	0.601 (0.159)	0.549 (0.152)	33.927	< .001	.113
Uber	0.340 (0.124)	0.433 (0.151)	0.359 (0.129)	0.322 (0.116)	27.741	< .001	.095
NDW-50	0.556 (0.162)	0.647 (0.156)	0.624 (0.176)	0.618 (0.183)	10.645	< .001	.039
VV1	0.525 (0.169)	0.540 (0.161)	0.536 (0.158)	0.484 (0.149)	5.188	.001	.019
VV2	0.451 (0.165)	0.467 (0.139)	0.506 (0.157)	0.477 (0.144)	4.649	.003	.017
SVV1	0.445 (0.181)	0.457 (0.166)	0.438 (0.153)	0.439 (0.171)	.504	.680	.002
AdvV	0.332 (0.182)	0.361 (0.191)	0.406 (0.192)	0.371 (0.173)	5.586	.001	.021
ModV	0.392 (0.159)	0.442 (0.161)	0.471 (0.165)	0.413 (0.146)	9.708	< .001	.035

Note: Values are means of the min-max-normalized index scores, with standard deviations in parentheses. The F statistics and effect sizes are invariant to this linear rescaling. Unadjusted omnibus p values are reported here; Holm-adjusted conventional and Welch results are provided in Appendix G. p values below .001 are reported as < .001.

Appendix D. Descriptive statistics and one-way ANOVA results for syntactic complexity indices across disciplines.

Index	Biology M (SD)	Chemistry M (SD)	Physics M (SD)	Engineering M (SD)	F(3, 796)	<i>p</i>	η^2
MLC	0.223 (0.139)	0.218 (0.089)	0.205 (0.106)	0.206 (0.111)	1.236	.295	.005
VP/T	0.299 (0.143)	0.279 (0.123)	0.312 (0.148)	0.318 (0.135)	3.159	.024	.012
C/T	0.422 (0.132)	0.415 (0.126)	0.433 (0.152)	0.413 (0.115)	.953	.415	.004
DC/C	0.409 (0.163)	0.390 (0.176)	0.413 (0.173)	0.405 (0.178)	.693	.557	.003
T/S	0.254 (0.084)	0.259 (0.106)	0.250 (0.091)	0.260 (0.099)	.462	.709	.002
CP/T	0.196 (0.124)	0.154 (0.070)	0.131 (0.071)	0.137 (0.071)	22.941	< .001	.080
CT/T	0.338 (0.159)	0.320 (0.164)	0.338 (0.173)	0.334 (0.170)	.524	.666	.002
CN/C	0.292 (0.166)	0.261 (0.113)	0.267 (0.128)	0.261 (0.132)	2.412	.066	.009

Note: Values are means of the min–max-normalized index scores, with standard deviations in parentheses. The F statistics and effect sizes are invariant to this linear rescaling. Unadjusted omnibus *p* values are reported here; Holm-adjusted conventional and Welch results are provided in Appendix G. *p* values below .001 are reported as < .001.

Appendix E. Tamhane's T2 / Tukey HSD post-hoc comparisons for lexical complexity indices.

Comparison	LD	LS1	LS2	VS1	MSTTR	Uber	NDW-50	VV1	VV2	SVV1	AdvV	ModV
B vs C	B < C*	B < C***	B < C***	B < C***	B < C***	B < C***	B < C***	-	-	-	-	B < C**
B vs E	-	-	-	-	B < E*	-	B < E**	B > E*	-	-	-	-
B vs P	-	B < p***	B < p***	-	B < p***	-	B < p***	-	B < p**	-	B < p***	B < p***
C vs E	-	C > E***	C > E***	C > E***	C > E***	C > E***	-	C > E**	-	-	-	-
C vs P	C > p**	C > p***	C > p***	C > p*	C > p**	C > p***	-	-	C < p*	-	-	-
E vs P	-	E < p*	-	E < p**	E < p**	E < p*	-	E < p**	-	-	-	E < p**

Note: B = Biology, C = Chemistry, E = Engineering, and P = Physics. Direction signs indicate which discipline had the higher mean. Tukey HSD was used when Levene's test was not significant; Tamhane's T2 was used when Levene's test was significant. - = no significant pairwise difference. * $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix F. Tamhane's T2 / Tukey HSD post-hoc comparisons for syntactic complexity indices.

Comparison	MLC	VP/T	C/T	DC/C	T/S	CP/T	CT/T	CN/C
B vs C	-	-	-	-	-	B > C***	-	-
B vs E	-	-	-	-	-	B > E***	-	-
B vs P	-	-	-	-	-	B > P***	-	-
C vs E	-	C < E*	-	-	-	-	-	-
C vs P	-	-	-	-	-	C > P**	-	-
E vs P	-	-	-	-	-	-	-	-

Note: B = Biology, C = Chemistry, E = Engineering, and P = Physics. Direction signs indicate which discipline had the higher mean. Tukey HSD was used when Levene's test was not significant; Tamhane's T2 was used when Levene's test was significant. - = no significant pairwise difference. * $p < .05$, ** $p < .01$, *** $p < .001$. Post hoc comparisons are not adjusted across indices; VP/T was not significant after Holm adjustment of the omnibus tests (Appendix G), so the C < E contrast on VP/T is not interpreted.

Appendix G. Assumption checks, robustness analyses, effect sizes, and Holm-adjusted results for the 20 retained indices.

Table G1a. Parametric robustness analyses for lexical indices

Index	Levene W	Levene p	F (3,796)	p	Holm p	η^2	95% CI	ω^2	Welch F	Welch df2	Welch p	Welch Holm p
LD	2.354	.071	5.158	.002	.016	0.019	[0.003, 0.039]	0.015	6.045	440.628	< .001	.006
LS1	11.616	< .001	38.839	< .001	< .001	0.128	[0.086, 0.169]	0.124	40.544	439.373	< .001	< .001
LS2	8.103	< .001	45.237	< .001	< .001	0.146	[0.102, 0.188]	0.142	43.452	439.642	< .001	< .001
VS1	0.358	.784	14.243	< .001	< .001	0.051	[0.023, 0.081]	0.047	13.868	442.159	< .001	< .001
NDW-50	1.089	.353	10.645	< .001	< .001	0.039	[0.015, 0.065]	0.035	11.814	441.428	< .001	< .001
MSTTR	2.077	.102	33.927	< .001	< .001	0.113	[0.073, 0.153]	0.110	37.018	441.477	< .001	< .001
Uber	2.158	.092	27.741	< .001	< .001	0.095	[0.057, 0.132]	0.091	23.951	440.554	< .001	< .001
VV1	1.479	.219	5.188	.001	.016	0.019	[0.003, 0.039]	0.015	5.645	441.844	< .001	.009
SVV1	1.768	.152	0.504	.680	1.000	0.002	[0.000, 0.008]	0.000	0.529	441.491	.662	1.000
VV2	1.895	.129	4.649	.003	.028	0.017	[0.002, 0.036]	0.013	4.234	441.374	.006	.052
AdvV	0.639	.590	5.586	< .001	.010	0.021	[0.004, 0.041]	0.017	5.419	441.889	.001	.011
ModV	0.552	.647	9.708	< .001	< .001	0.035	[0.012, 0.061]	0.032	9.319	441.753	< .001	< .001

Table G1b. Parametric robustness analyses for syntactic indices

Index	Levene W	Levene p	F(3, 796)	p	Holm p	η^2	95% CI	ω^2	Welch F	Welch df2	Welch p	Welch Holm p
MLC	2.787	.040	1.236	.295	1.000	0.005	[0.000, 0.015]	0.001	1.154	437.699	.327	1.000
C/T	2.131	.095	0.953	.415	1.000	0.004	[0.000, 0.013]	0.000	0.853	440.406	.466	1.000
CT/T	0.992	.396	0.524	.666	1.000	0.002	[0.000, 0.009]	0.000	0.535	442.026	.659	1.000
DC/C	0.592	.621	0.693	.557	1.000	0.003	[0.000, 0.010]	0.000	0.678	441.965	.566	1.000
CP/T	11.231	< .001	22.941	< .001	< .001	0.080	[0.045, 0.115]	0.076	15.852	436.202	< .001	< .001
T/S	0.452	.716	0.462	.709	1.000	0.002	[0.000, 0.008]	0.000	0.456	440.686	.713	1.000
CN/C	4.549	.004	2.412	.066	.459	0.009	[0.000, 0.023]	0.005	1.870	439.085	.134	.937
VP/T	0.794	.498	3.159	.024	.193	0.012	[0.000, 0.028]	0.008	3.569	441.206	.014	.113

Note: The numerator degrees of freedom for Welch tests are 3 for every index. Confidence intervals refer to η^2 . Values are rounded to three decimal places; p values below .001 are shown as < .001.

Note: N = 800 (200 abstracts per discipline). Levene's tests were median-centered. Holm adjustment was applied separately across the 20 conventional ANOVA tests and the 20 Welch ANOVA tests. η^2 equals η^2 in this one-factor design; 95% confidence intervals were obtained from the noncentral F distribution. Negative sampling estimates of ω^2 were reported as 0.

Distributional diagnostics and rank-based robustness. Residual Q–Q plots and discipline-specific boxplots showed noticeable right-tail departures and isolated extreme observations for several syntactic ratio measures, especially MLC, CP/T, T/S, CN/C, and VP/T. Holm-adjusted Kruskal–Wallis tests reproduced the main dimensional pattern: all lexical indices except SVV1 were significant, whereas CP/T was the only significant syntactic index. VV2 remained significant in this rank-based analysis (Holm p = .026) but was borderline in the Welch–Holm analysis (p = .052), so it is treated as procedure-sensitive rather than uniformly robust.

Table G2. Rank-based robustness analyses for the 20 retained indices

Index	Kruskal–Wallis H(3)	p	Holm-adjusted p	Interpretation
LD	16.071	.001	.011	Disciplinary difference retained
LS1	97.405	< .001	< .001	Disciplinary difference retained
LS2	117.690	< .001	< .001	Disciplinary difference retained
VS1	42.634	< .001	< .001	Disciplinary difference retained
MSTTR	90.511	< .001	< .001	Disciplinary difference retained
Uber	69.340	< .001	< .001	Disciplinary difference retained
NDW-50	33.745	< .001	< .001	Disciplinary difference retained
VV1	17.453	< .001	.007	Disciplinary difference retained
VV2	14.009	.003	.026	Disciplinary difference retained
SVV1	2.028	.567	1.000	Not retained after Holm
AdvV	16.498	< .001	.010	Disciplinary difference retained
ModV	30.878	< .001	< .001	Disciplinary difference retained

Index	Kruskal–Wallis H(3)	p	Holm-adjusted p	Interpretation
MLC	4.136	.247	1.000	Not retained after Holm
C/T	0.823	.844	1.000	Not retained after Holm
CT/T	1.579	.664	1.000	Not retained after Holm
DC/C	2.911	.406	1.000	Not retained after Holm
CP/T	53.384	< .001	< .001	Disciplinary difference retained
T/S	1.739	.628	1.000	Not retained after Holm
CN/C	2.567	.463	1.000	Not retained after Holm
VP/T	9.537	.023	.183	Not retained after Holm

Note. Holm adjustment was applied across the family of 20 Kruskal–Wallis tests. N = 800 (200 abstracts per discipline).

Appendix H. Omnibus non-inferiority tests for the eight syntactic indices.

Table H1. Omnibus non-inferiority tests at the post-specified bound $\Delta = .02$

Index	F(3, 796)	η^2	90% CI	Non-inferiority p	Holm p	Decision
MLC	1.236	0.005	[0.000, 0.012]	.007	.027	Effect below 0.02
C/T	0.953	0.004	[0.000, 0.010]	.003	.016	Effect below 0.02
CT/T	0.524	0.002	[0.000, 0.006]	< .001	.004	Effect below 0.02
DC/C	0.693	0.003	[0.000, 0.008]	.001	.008	Effect below 0.02
CP/T	22.941	0.080	[0.050, 0.108]	1.000	1.000	Inconclusive / not below bound
T/S	0.462	0.002	[0.000, 0.006]	< .001	.004	Effect below 0.02
CN/C	2.412	0.009	[0.000, 0.020]	.049	.147	Inconclusive / not below bound
VP/T	3.159	0.012	[0.001, 0.024]	.106	.213	Inconclusive / not below bound

Note. The omnibus tests evaluated $H_0: \eta^2 \geq \Delta$ against $H_1: \eta^2 < \Delta$ for the overall four-discipline effect (N = 800), using the conventional one-way ANOVA F statistic and a boundary noncentrality parameter of $N\Delta/(1 - \Delta)$ (Campbell & Lakens, 2021). Holm adjustment was applied across all eight syntactic indices at each bound. The primary $\Delta = .02$ bound was post-specified and lies between conventional small (.01) and medium (.06) reference values; the results are sensitivity analyses rather than confirmatory evidence of complete syntactic equivalence. Because conventional F statistics were used for all eight indices, conclusions for heteroscedastic MLC, CN/C, and CP/T should be read cautiously.

Boundary sensitivity of the omnibus non-inferiority conclusions

Bound (Δ)	Indices below the bound after Holm adjustment	Indices not shown to be below the bound
.01	None	MLC, C/T, CT/T, DC/C, CP/T, T/S, CN/C, VP/T
.02	MLC, C/T, CT/T, DC/C, T/S	CP/T, CN/C, VP/T
.03	MLC, C/T, CT/T, DC/C, T/S, CN/C, VP/T	CP/T

Appendix I. Comparison of the EWM-weighted and unweighted composite scores for the final 20 indices.

Composite (final 20 indices)	Disciplinary ordering	F(3, 796)	p	η^2
EWM-weighted LC score	Chemistry > Physics > Biology > Engineering	7.382	< .001	.027
Unweighted mean	Chemistry > Physics > Engineering > Biology	18.699	< .001	.066

Note. Both composites were computed from the same 20 min–max-normalized indices. The two composites correlated strongly (Pearson $r = .945$; Spearman $\rho = .936$; both $p < .001$). Disciplinary effects are from one-way ANOVAs on $N = 800$ texts.

References

- Ädel, A. (2014). Selecting quantitative data for qualitative analysis: A case study connecting a lexicogrammatical pattern to rhetorical moves. *Journal of English for Academic Purposes*, 16, 68–80. <https://doi.org/10.1016/j.jeap.2014.09.001>
- Anjarwati, D., Basthomi, Y., & Zen, E. L. (2024). Comparative analysis of syntactic complexity in Indonesian graduate-level writing: A corpus study of master's theses versus doctoral dissertations. *English Review: Journal of English Education*, 12(2), 811–820. <https://doi.org/10.25134/erjee.v12i2.9528>
- Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes*, 31, 58–71. <https://doi.org/10.1016/j.jeap.2017.12.008>
- Ayadi, M. (2024). Exploring lexical complexity in scientific writing: A corpus-based study across disciplines. *LLT Journal: A Journal on Language and Language Teaching*, 27(1), 421–433. <https://doi.org/10.24071/llt.v27i1.6790>
- Bao, K., & Liu, M. (2023). Comparative analysis of move-specific lexical bundles in linguistics dissertation abstracts: A study of students from China and the United States. *SAGE Open*, 13(4), Article 21582440231210055. <https://doi.org/10.1177/21582440231210055>
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20. <https://doi.org/10.1016/j.jeap.2010.01.001>
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5–35. <https://doi.org/10.5054/tq.2011.244483>
- Biber, D., Gray, B., Staples, S., & Egbert, J. (2021). *The register-functional approach to grammatical complexity: Theoretical foundation, descriptive research findings, application*. Routledge. <https://doi.org/10.4324/9781003087991>
- Bui, G. (2021). Influence of learners' prior knowledge, L2 proficiency, and pre-task planning on L2 lexical complexity. *International Review of Applied Linguistics in Language Teaching*, 59(4), 543–567. <https://doi.org/10.1515/iral-2018-0244>
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In I. Vedder, A. Housen, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency* (pp. 21–46). John Benjamins. <https://doi.org/10.1075/llt.32.02bul>
- Burrough-Boenisch, J. (2003). Shapers of published NNS research articles. *Journal of Second Language Writing*, 12(3), 223–243. [https://doi.org/10.1016/S1060-3743\(03\)00037-7](https://doi.org/10.1016/S1060-3743(03)00037-7)
- Campbell, H., & Lakens, D. (2021). Can we disregard the whole model? Omnibus non-inferiority testing for R^2 in multiple regression and η^2 in ANOVA. *British Journal of Mathematical and Statistical Psychology*, 74(1), 64–89. <https://doi.org/10.1111/bmsp.12201>
- Casal, J. E., Lu, X., Qiu, X., Wang, Y., & Zhang, G. (2021). Syntactic complexity across academic research article part-genres: A cross-disciplinary perspective. *Journal of English for Academic Purposes*, 52, Article 100996. <https://doi.org/10.1016/j.jeap.2021.100996>

- Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity and errors. *System*, 96, Article 102408. <https://doi.org/10.1016/j.system.2020.102408>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. *International Review of Social Psychology*, 32(1), Article 13. <https://doi.org/10.5334/irsp.198>
- Dong, J., Wang, H., & Buckingham, L. (2023). Mapping out the disciplinary variation of syntactic complexity in student academic writing. *System*, 113, Article 102974. <https://doi.org/10.1016/j.system.2022.102974>
- Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165–193. <https://doi.org/10.1093/applin/amv011>
- El-Dakhs, D. A. S. (2018). Why are abstracts in PhD theses and research articles different? A genre-specific perspective. *Journal of English for Academic Purposes*, 36, 48–60. <https://doi.org/10.1016/j.jeap.2018.09.005>
- Hao, J., Jin, Q., & Gao, Y. (2026). Knowledge construction modes in humanities and social sciences: A cross-disciplinary analysis through linguistic complexity lens. *Journal of English for Academic Purposes*, 79, Article 101600. <https://doi.org/10.1016/j.jeap.2025.101600>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Hu, G., & Bonsu, E. M. (2025). A cross-disciplinary study of value arguments in doctoral theses submitted to universities in Hong Kong. *English for Specific Purposes*, 79, 1–16. <https://doi.org/10.1016/j.esp.2025.02.002>
- Huang, Y., Li, D., & Cheung, A. K. F. (2025). Evaluating the linguistic complexity of machine translation and LLMs for EFL/ESL applications: An entropy weight method. *Research Methods in Applied Linguistics*, 4(3), Article 100229. <https://doi.org/10.1016/j.rmal.2025.100229>
- Hyland, K. (2002). Specificity revisited: How far should we go now? *English for Specific Purposes*, 21(4), 385–395. [https://doi.org/10.1016/S0889-4906\(01\)00028-X](https://doi.org/10.1016/S0889-4906(01)00028-X)
- Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing* (Michigan Classics ed.). University of Michigan Press. <https://doi.org/10.3998/mpub.6719>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K. (2015). *Academic publishing: Issues and challenges in the construction of knowledge*. Oxford University Press.
- Hyland, K., & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly*, 41(2), 235–253. <https://doi.org/10.1002/j.1545-7249.2007.tb00058.x>
- Jiang, F. (Kevin), & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes*, 46, 1–14. <https://doi.org/10.1016/j.esp.2016.11.001>
- Khany, R., & Kafshgar, N. B. (2016). Analysing texts through their linguistic properties: A cross-disciplinary study. *Journal of Quantitative Linguistics*, 23(3), 278–294. <https://doi.org/10.1080/09296174.2016.1169848>
- Kim, M. (2021). Exploring longitudinal changes in lexical and syntactic features in beginning-level EFL learner writing. *System*, 103, Article 102680. <https://doi.org/10.1016/j.system.2021.102680>

- Kim, Y. L., Crawford, W. J., & McDonough, K. (2025). The role of communicative purpose in describing and interpreting lexico-grammatical variation in L2 writing. *Journal of Second Language Writing*, 67, Article 101176. <https://doi.org/10.1016/j.jslw.2025.101176>
- Koutsantoni, D. (2006). Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes*, 5(1), 19–36. <https://doi.org/10.1016/j.jeap.2005.11.002>
- Kwok, H. L., Shi, Y., Xu, H., Li, D., & Liu, K. (2025). Gen-AI as a translation assistant? A corpus-based study on lexical and syntactic complexity of GPT-post-edited learner translation. *System*, 130, 103618. <https://doi.org/10.1016/j.system.2025.103618>
- Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333–349. <https://doi.org/10.1111/modl.12468>
- Larsson, T., & Hancock, G. R. (2026). What if we want to test for similarities between groups, rather than differences? Equivalence testing techniques for corpus linguistics. *Corpora*, 21(1), 21–33. <https://doi.org/10.3366/cor.2026.0353>
- Lei, L. (2017). Zhongguo Yingyu xuexizhe xueshu xiezuo jufa fuzadu yanjiu [A study of syntactic complexity in academic writing by Chinese EFL learners]. *Journal of PLA University of Foreign Languages*, 40(5), 1–10, 159.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190–208. <https://doi.org/10.1111/j.1540-4781.2011.01232.x>
- Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, 39, 21–36. <https://doi.org/10.1016/j.jeap.2019.03.008>
- Lu, X., Casal, J. E., Liu, Y., Kisselev, O., & Yoon, J. (2021). The relationship between syntactic complexity and rhetorical move-steps in research article introductions: Variation among four social science and engineering disciplines. *Journal of English for Academic Purposes*, 52, Article 101006. <https://doi.org/10.1016/j.jeap.2021.101006>
- Luo, N., & Hyland, K. (2016). Chinese academics writing for publication: English teachers as text mediators. *Journal of Second Language Writing*, 33, 43–55. <https://doi.org/10.1016/j.jslw.2016.06.005>
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, Article 100511. <https://doi.org/10.1016/j.asw.2020.100511>
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578. <https://doi.org/10.1093/applin/amp044>
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Paltridge, B., & Starfield, S. (2020). *Thesis and dissertation writing in a second language: A handbook for students and their supervisors* (2nd ed.). Routledge.
- Pan, F., & Yang, Y. (2025). Diachronic change in lexical complexity of research articles (1970–2020): Economics vs. medicine. *Scientometrics*, 130(3), 1789–1812. <https://doi.org/10.1007/s11192-025-05258-6>
- Pérez-Guerra, J., & Smirnova, E. A. (2023). How complex is professional academic writing? A corpus-based analysis of research articles in “hard” and “soft” disciplines. *Vigo International Journal of Applied Linguistics*, 20(20), 149–183. <https://doi.org/10.35869/vial.v0i20.4357>

- Picazo-Sanchez, P., & Ortiz-Martin, L. (2024). Analysing the impact of ChatGPT in research. *Applied Intelligence*, 54(5), 4172–4188. <https://doi.org/10.1007/s10489-024-05298-0>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.
- Shen, C., Guo, J., Shi, P., Qu, S., & Tian, J. (2023). A corpus-based comparison of syntactic complexity in academic writing of L1 and L2 English students across years and disciplines. *PLOS ONE*, 18(10), Article e0292688. <https://doi.org/10.1371/journal.pone.0292688>
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, 33(2), 149–183. <https://doi.org/10.1177/0741088316631527>
- Storer, N. W. (1967). The hard sciences and the soft: Some sociological observations. *Bulletin of the Medical Library Association*, 55(1), 75–84.
- Swales, J. M., & Feak, C. B. (2009). *Abstracts and the writing of abstracts*. University of Michigan Press. <https://doi.org/10.3998/mpub.309332>
- Tan, L. (2024). *NeoSCA* (Version 0.1.4) [Computer software]. GitHub. <https://github.com/tanloong/neosca>
- Urbaniak, G. C., & Plous, S. (2013). *Research Randomizer* (Version 4.0) [Computer software]. <https://www.randomizer.org/>
- Wang, M., & Lowie, W. (2021). Understanding advanced level academic writing on syntactic complexity. In K. Hu, J.-B. Kim, C. Zong, & E. Chersoni (Eds.), *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation* (pp. 452–462). Association for Computational Linguistics. <https://aclanthology.org/2021.paclic-1.48/>
- Wang, Q., Yuan, X., Zhang, J., Gao, Y., Hong, J., Zuo, J., & Liu, W. (2015). Assessment of the sustainable development capacity with the entropy weight coefficient method. *Sustainability*, 7(10), 13542–13563. <https://doi.org/10.3390/su71013542>
- Wong, W. L., & Fong, S. (2025). A new era with artificial intelligence powered learner corpora for teaching and learning of English as a second language in primary school education. *Applied Language Sciences*, 1, 250105. <https://doi.org/10.65553/ALS.250105>
- Xiao, W., & Sun, S. (2020). Dynamic lexical features of PhD theses across disciplines: A text mining approach. *Journal of Quantitative Linguistics*, 27(2), 114–133. <https://doi.org/10.1080/09296174.2018.1531618>
- Yang, J., Lan, G., & Huang, M. (2026). Disciplinary variation in grammatical complexity in university science textbooks in Hong Kong. *Journal of English for Academic Purposes*, 79, Article 101624. <https://doi.org/10.1016/j.jeap.2025.101624>
- Yasuda, S. (2024). Does “more complexity” equal “better writing”? Investigating the relationship between form-based complexity and meaning-based complexity in high school EFL learners’ argumentative writing. *Assessing Writing*, 61, Article 100867. <https://doi.org/10.1016/j.asw.2024.100867>
- Yi, L., & He, Z. (2026). Syntactic complexity in knowledge construction across disciplines: Evidence from dependency distance. *Journal of English for Academic Purposes*, 79, Article 101617. <https://doi.org/10.1016/j.jeap.2025.101617>
- Zhang, X., & Lu, X. (2024). Testing the relationship of linguistic complexity to second language learners’ comparative judgment on text difficulty. *Language Learning*, 74(3), 672–706. <https://doi.org/10.1111/lang.12633>
- Zhao, W. (2025). Reconstructing stance in EFL doctoral thesis writing through generative artificial intelligence. *Humanities and Social Sciences Communications*, 12(1), Article 1963. <https://doi.org/10.1057/s41599-025-06249-x>

- Zhou, W., Li, Z., & Lu, X. (2023). Syntactic complexity features of science research article introductions: Rhetorical-functional and disciplinary variation perspectives. *Journal of English for Academic Purposes*, 61, Article 101212. <https://doi.org/10.1016/j.jeap.2022.101212>
- Zhu, H., Wang, T., & Pang, N. (2024). Diachronic changes in lexical density of research article abstracts: A corpus-based study. *Lingua*, 312, Article 103837. <https://doi.org/10.1016/j.lingua.2024.103837>
- Ziaeeian, E., & Golparvar, S. E. (2022). Fine-grained measures of syntactic complexity in the discussion section of research articles: The effect of discipline and language background. *Journal of English for Academic Purposes*, 57, Article 101116. <https://doi.org/10.1016/j.jeap.2022.101116>
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>

Ying Chai, Ph.D. candidate, Department of Language Science and Technology, The Hong Kong Polytechnic University. ORCID: 0009-0003-2811-0914

Yingqi Huang, Ph.D. candidate, Department of Language Science and Technology, The Hong Kong Polytechnic University. ORCID: 0009-0000-1567-5578

Yueyue Huang, Associate Professor, College of Foreign Languages, Fujian Normal University, Fuzhou, Fujian, China. ORCID: 0000-0003-1864-0367

Jing Deng, Lecturer, Division of Languages and Communication, College of Professional and Continuing Education, The Hong Kong Polytechnic University. ORCID: 0000-0001-8606-1661