

A Corpus-Based Contrastive Study of Noun Phrase Complexity in Rhetorical Move Structures of Sociology Abstracts by L1 Chinese and L1 English Authors

Qiaoqiao Chen¹, Zhijun Li^{2*}

¹² Huaqiao University, Fujian Province, China

Abstract

This study compares rhetorical move structures and noun phrasal complexity patterns embedded therein in two corpora, each containing 160 sociology research article abstracts written by L1 Chinese and L1 English writers. A corpus-based comparative approach was adopted to examine move structures and noun phrasal complexity, with Mann-Whitney U tests and Bonferroni correction used to identify differences in noun phrasal complexity between the two corpora. The findings show that both groups follow the conventional five-move structure, but differ in the distribution of individual moves. L1 English writers tend to establish research gaps, whereas L1 Chinese writers place more emphasis on research contributions. Noun phrasal complexity also varies systematically across moves to serve different communicative functions. L1 Chinese writers rely more on phrasal compression through attributive adjectives, pre-modifying nouns, and prepositional phrases (of), particularly in background- and conclusion-oriented moves. In contrast, L1 English writers use more clausal post-modification, including finite relative and complement clauses, especially in result-oriented moves. Implications for multilingual academic writing, EFL writing, and EAP pedagogy are discussed.

KEYWORDS

sociology, research article abstracts, rhetorical move structure, noun phrasal complexity, EFL writing

Introduction

For the past few decades, phrasal complexity has become an important construct in academic writing research, providing insights into L2 learners' development and writing proficiency (Lan et al., 2022). As a sub-construct of syntactic complexity, it concerns grammatical devices such as phrasal elaboration (Shao et al., 2022) and has been widely recognized as an indicator of writing quality in academic discourse (Thongyoi & Poonpon, 2020). In particular, noun phrasal complexity is crucial in academic writing because it enables writers to condense disciplinary information into concise linguistic forms. This function is especially relevant to research article (hereafter RA) abstracts, which serve as stand-alone "mini-texts" that summarize research, support indexing, and help readers assess article relevance (Huckin, 2001). With the increasing importance of international academic communication, producing rhetorically effective and linguistically appropriate English abstracts has become essential for academic writers worldwide (Swales, 1990). Therefore, the ability to produce rhetorically effective and linguistically sophisticated abstracts is increasingly important for academic writers worldwide.

A growing body of research has demonstrated that phrasal complexity in academic writing is influenced by multiple factors, including language backgrounds

(Lan et al., 2022), genre conventions (Y. X. Wang & Jiang, 2024), and academic levels (Ansarifar et al., 2018). Previous studies have further examined phrasal complexity across rhetorical moves and its relationship with disciplinary writing practices (Parviz & Lan, 2023), suggesting that linguistic complexity is systematically associated with the communicative purposes of different moves rather than functioning as an isolated textual feature.

Despite these insightful studies, several limitations remain. First, although previous research has examined rhetorical move structures and syntactic complexity separately, few studies have explored noun phrasal complexity across rhetorical moves in RA abstracts. Second, while academic writing has been shown to vary with writers' syntactic preferences (Shen et al., 2023), nominalization practices (Terblanche, 2009), and patterns of academic socialization (Leki et al., 2008), limited attention has been paid to how writers from different linguistic and academic backgrounds, particularly L1 Chinese and L1 English writers, employ noun phrasal resources across rhetorical moves. Therefore, this study compares English RA abstracts by L1 Chinese and L1 English writers within the same discipline to examine whether noun phrasal complexity patterns associated with specific rhetorical moves are consistently realized across writer groups.

Against this backdrop, the present study examines noun phrasal complexity across rhetorical moves in sociology RA abstracts written by Chinese and English scholars. Specifically, it addresses the following research questions:

- (1) What similarities and differences are there in the rhetorical move structures of sociology RA abstracts authored by L1 Chinese and L1 English writers?
- (2) What similarities and differences exist in the distribution of noun phrasal complexity across different rhetorical moves in sociology RA abstracts authored by L1 Chinese and L1 English writers?

By examining how writers with different L1 backgrounds employ noun phrasal resources to realize rhetorical functions, this study enhances understanding of the relationship between rhetorical move structure and linguistic complexity in sociology RA abstracts. It also offers pedagogical implications by emphasizing the alignment of linguistic choices with disciplinary rhetorical expectations in English academic writing.

Literature Review

Studies on rhetorical move structures of RA abstracts

In academic prose, abstracts, as "a part-genre of the RA" (Swales & Feak, 2009), provide selective representations of research (Hyland, 2000), enabling readers to access key information and authors present their research effectively in academic publishing contexts (Bondi, 2014). As an independent section of RAs, abstracts are organized through rhetorical structures that reflect their communicative purposes.

In genre-based discourse analysis, a move is generally regarded as a functional or rhetorical unit that serves a specific communicative purpose (Swales, 2004). Genres are constructed through patterned sequences of moves, and the combination of these moves contributes to the realization of overall communicative goals (Swales, 1990). Based on this perspective, various move frameworks have been proposed for RA abstracts, including the models developed by Bhatia (1993), Santos (1996), and Hyland (2000).

Empirical studies have demonstrated disciplinary

variation in abstract move structures across fields, including applied linguistics (Pho, 2008; Santos, 1996; Tseng, 2011), engineering (Lam & Tan, 2012; Maswana et al., 2015), literature (Tankó, 2017), chemistry (Darabad, 2016), and sociology (Z. J. Li et al., 2025; Lorés-Sanz, 2014). Previous studies have also reported variations in abstract move structures across writers with different linguistic backgrounds (Busch-Lauer, 1995). For example, Tseng (2011) identified differences between native and non-native English writers in applied linguistics abstracts, while Busch-Lauer (1995) found that German writers differed from native English writers in the use of background and purpose moves.

Given that writers with different L1 backgrounds may engage with academic discourse conventions differently, such differences may be reflected in abstract move structures. However, despite extensive research on RA abstract move structures, limited attention has been paid to how writers with different L1 backgrounds organize these structures. In particular, little is known about differences in the rhetorical move structures of sociology RA abstracts written by L1 Chinese and L1 English writers. Therefore, this study attempts to fill this gap.

Studies on phrasal complexity

Complexity in academic writing has been closely associated with the ability to condense information through elaborated nominal structures, alongside a reduced reliance on clausal elaboration (Biber & Gray, 2016). From a systemic functional perspective, noun phrases provide important resources for packaging complex meanings through nominalization, premodification, and post-modification (Halliday & Matthiessen, 2013). Accordingly, recent studies have shifted from traditional clausal measures toward fine-grained analyses of noun phrasal modification, including attributive adjectives, noun premodifiers, prepositional phrases, participle clauses, and appositive structures (Biber et al., 2011; Pan & Yang, 2024). These NP-based measures offer a more functionally interpretable account of advanced academic writing than conventional syntactic indices (Kyle & Crossley, 2018; Lu, 2011).

Previous research has shown that noun phrasal complexity varies with both writers' proficiency and disciplinary conventions. Higher-proficiency writers tend to employ more diverse and elaborated modification structures to achieve greater informational density (Ansarifar et al., 2018; Lan et al., 2019; Parkinson & Musgrave, 2014). Meanwhile, noun phrasal modification patterns vary across disciplines. Hard-science writing generally tends to favor dense nominal premodification, whereas humanities and social science writing generally relies more on post-modifying structures to support interpretation and argumentation (Hu & He, 2023). Corpus-based studies have further confirmed disciplinary variation in noun phrase complexity across academic fields (Pérez-Guerra & Smirnova, 2023).

More recently, research has extended attention to RA abstracts, a compressed genre requiring writers to fulfill multiple rhetorical purposes within limited textual space. Such communicative demands encourage intensive information packaging and frequent use of complex noun phrases. Ruan (2018) found that writers with different L1 backgrounds showed variation in the realization of phrasal structures in English RA abstracts, while (Y. Wang et al., 2022) reported disciplinary differences in abstract writing, with hard-science abstracts favoring nominal premodification and soft-science abstracts employing more clausal postmodification. These findings suggest that noun phrase complexity in abstracts may be

associated with both genre demands and disciplinary conventions.

Despite these advances, research on noun phrasal complexity in RA abstracts remains limited, particularly in sociology. Existing research has mainly relied on mixed-disciplinary corpora or broad disciplinary comparisons, with discipline-specific patterns remaining insufficiently examined. Although disciplinary and L1 differences have been examined separately, few studies have investigated noun phrase complexity in sociology RA abstracts across L1 backgrounds. Consequently, it remains unclear whether L1 Chinese and L1 English writers employ similar or different noun phrasal modification patterns in this highly compressed genre. Therefore, there is a need for a discipline-specific comparison of noun phrasal complexity in sociology RA abstracts written by L1 Chinese and L1 English writers to examine variation in their noun phrase modification patterns.

Studies on the combination of linguistic complexity and rhetorical structures

Earlier research on academic discourse often examined rhetorical move structures (Sükan & Mohammadzadeh, 2022) or linguistic features such as hedging (Xie & Mi, 2023) and lexical bundles (Gong et al., 2025) separately. Recent studies, however, have increasingly emphasized the relationship between rhetorical and linguistic dimensions by exploring how linguistic choices realize communicative functions (Hyland, 2004).

A growing body of research has demonstrated that rhetorical moves are associated with specific linguistic realizations. Studies on lexical bundles, clause structures, and phrase patterns across move functions have shown that different moves employ different linguistic resources to achieve their communicative purposes (Cortés, 2013; L. J. Li & Ge, 2009). More recent research has extended this perspective to linguistic complexity, including syntactic complexity (Lu et al., 2020), phrasal complexity (Parviz & Lan, 2023), and nominalization (Arizavi et al., 2015), suggesting that complexity patterns vary across rhetorical functions. For example, Casal et al., (2021) and Parviz and Lan (2023) found that syntactic and phrasal complexity vary across move-steps, while Arizavi et al. (2015) showed that nominalization supports information condensation in advanced rhetorical stages. These findings are consistent with corpus-based research showing that academic writing relies on NP-based compression through extensive modification and dense nominal structure (Biber & Gray, 2011, 2016). Such nominal resources facilitate information packaging and disciplinary meaning construction, highlighting the role of noun phrase complexity in realizing rhetorical functions.

However, there are still three main research gaps in the aforementioned research. First, research on rhetorical move structures has predominantly focused on hard-science disciplines, while sociology, as a soft science characterized by interpretation and conceptual reasoning rather than experimental verification (Becher & Trowler, 2001), remains underexplored. Therefore, it is unclear whether patterns identified in scientific disciplines apply to the sociological discourse community. Second, existing contrastive studies have often examined rhetorical structures and linguistic features separately, with limited attention to their interaction within comparable datasets. Third, comparative studies involving writers from different L1

backgrounds often treat these groups as relatively homogeneous, overlooking internal variation (Cumming, 2013; Kobayashi & Rinnert, 2008). However, differences in L1 background may influence writers' rhetorical strategies and noun phrasal complexity, but their role in shaping variation in academic writing remains underexplored.

In response to these issues, the present study compares rhetorical move structures and noun phrasal complexity embedded therein in sociology RA abstracts written by L1 Chinese and L1 English writers. By examining how noun phrasal resources are distributed across rhetorical moves to achieve communicative purposes, this study contributes to a more nuanced understanding of the interaction between rhetorical purposes and linguistic complexity in sociology academic writing.

Methods

Corpus

The data for this study are two self-built corpora, each comprising 160 English abstracts of empirical RAs published in Chinese and international sociology journals in 2021-2024: the Chinese Academic Abstracts Corpus (CAAC) and the English Academic Abstracts Corpus (EAAC).

To ensure corpus comparability, journals were selected based on the Social Sciences Citation Index (SSCI) and the Chinese Social Sciences Citation Index (CSSCI). The EAAC comprises abstracts from four SSCI-indexed sociology journals, namely American Journal of Sociology (AJS), American Sociological Review (ASR), Social Forces (SF), and The British Journal of Sociology (BJS). The CAAC includes abstracts from four CSSCI-indexed sociology journals in Mainland China, namely Sociological Research (SR), Society (S), Journal of Social Development (JSD), and Population Research (PR). These journals publish empirical sociological research and reflect major publication practices in their respective sociological contexts. The publication period was restricted to 2021-2024 to minimize diachronic variation and reflect current practices in sociology research writing.

Authors were classified into L1 Chinese and L1 English groups based on established criteria in corpus-based academic writing research. This classification was intended to examine variation associated with writers' L1 backgrounds, rather than language differences, as all abstracts analyzed were written in English L1. L1 Chinese writers were identified based on Chinese names, education at Mainland Chinese universities, and affiliations with Mainland Chinese institutions, whereas L1 English writers were identified based on English names, education in English-speaking countries (e.g., the US, UK, Canada, and Australia), and corresponding affiliations. To reduce classification ambiguity, only articles with no more than three co-authors were included, and classifications were determined using multiple sources of evidence.

A stratified random sampling strategy was adopted to ensure proportional representation across journals. Only empirical RA abstracts were included, while non-empirical genres, such as book reviews, editorials, special-issue introductions, and theoretical essays, were excluded. Abstracts were extracted based on explicit structural markers (e.g., "Abstract"), and data cleaning involved removing duplicate authors and abstracts shorter than 60 words to ensure comparability between the two corpora. The two corpora total 61011 words. Detailed information is

provided in [Table 1](#).

Table 1. Basic information of the two corpora

Corpus	N	Words	Mean of length (words)	Std. Deviation
EAAC	160	29342	183.4	34.52
CAAC	160	31669	197.9	67.83

Analytical framework

To analyze the rhetorical move structure of abstracts, Li et al.'s (2025) framework for sociology abstracts was adopted (see [Table 2](#)). Developed specifically for sociology RAs, this framework provides a detailed scheme for move identification and includes strategic sub-steps within Move 1 and Move 5. Structurally, it comprises five rhetorical moves, with Move 1 and Move 5 further subdivided into functional steps, while Moves 2, 3, and 4 operate as single units. This fine-grained classification enables a more precise examination of the distribution of noun phrasal complexity across rhetorical functions.

The analysis of noun phrasal complexity was based on [Pan's \(2024\)](#) framework (see [Table 3](#)), which was chosen for three reasons. First, its fine-grained classification captures subtle structural differences among noun phrases, enabling comparisons between L1 Chinese and L1 English writers and linking noun phrasal complexity to rhetorical functions. Second, its increasing application in recent RA studies ([Pan & Yang, 2024](#)) demonstrates its applicability and facilitates comparison with previous research. Third, its operationalized criteria ensure a consistent and reliable analytical process. To further refine this framework, the present study extends [Pan's \(2024\)](#) model by incorporating appositive noun phrases which represent an important indicator of phrasal complexity and reflect the tendency toward grammatical compression and informational density in academic writing. ([Biber et al.,](#)

[1999; Biber & Gray, 2016](#)). In the extended framework, appositive structures are categorized as "Dependent phrases" of the "Phrasal" type and placed in the "Post" position.

Annotation procedure

The annotation procedure consisted of two stages: annotation of rhetorical moves/steps and that of noun phrasal complexity. Rhetorical moves/steps were first manually annotated using UAM CorpusTool 2.0 (UAMCT 2.0). Each abstract was segmented and annotated to identify move boundaries and communicative functions, providing the basis for analyzing noun phrasal complexity across move-step units. Subsequently, noun phrasal complexity features were identified through a hybrid procedure. Python-based scripts were developed to automatic retrieval and the extracted results were manually checked and corrected for parsing and identification errors.

To ensure annotation reliability, the coding criteria were discussed by the two authors prior to annotation. Afterwards, the first author conducted the initial and final annotations approximately one month apart, with ambiguous cases reviewed according to the coding criteria and discussed with the second author when necessary. The intra-coder agreement rate reached 95.6% for move-step identification and 92.1% for phrasal complexity annotation. After annotations, frequency data for rhetorical move-step distributions and noun phrasal complexity features were extracted for statistical analyses.

The normality of noun phrasal complexity indices was assessed using the Shapiro-Wilk test. As all indices deviated from normality, Mann-Whitney U tests were conducted to examine differences between L1 Chinese and L1 English writers across rhetorical move/steps. The significance level was set at .05, with Bonferroni correction applied to control Type I error from multiple comparisons, resulting in an adjusted threshold of $p < .00556$ (0.05/9).

Table 2. Framework for the rhetorical move structure of abstracts ([Li et al., 2025](#))

Moves/Step	Tag	Description
Move 1 (Background)	M1	Introducing the study within its field
Step1	M1_S1	Providing background information
Step2	M1_S2	Claiming importance of the research topic
Step3	M1_S3	Indicating a gap in existing literature
Move 2 (Topic)	M2	Stating the aim / topic / hypothesis of the study
Move 3 (Method)	M3	Describing the research method (the materials, subjects, method, etc.)
Move 4 (Results)	M4	Reporting main results
Move 5 (Conclusion)	M5	Interpreting results and/or drawing conclusions / implications
Step1	M5_S1	Interpreting or explaining results
Step2	M5_S2	Drawing conclusions from results
Step3	M5_S3	Stating the significance or contribution of the research findings

Table 3. Noun phrasal complexity measures adapted from [Pan's \(2024\)](#) Framework

Structural types	Grammatical subcategories	Type	Position	Examples
Finite dependent clauses	Finite relative clauses (FRC)	Clausal	Post	experience <i>where precarity originates</i>
	Noun + <i>that</i> complement clauses (NThatCC)	Clausal	Post	the evidence <i>that wealth is a distinct dimension of SES</i>
Non-finite dependent clauses	- <i>ing</i> Non-finite relative clauses (-ing NFRC)	Clausal	Post	Populations <i>living within a state</i>
	- <i>ed</i> Non-finite relative clauses (-ed NFRC)	Clausal	Post	Study <i>conducted in a representative German online</i>
	Noun + <i>to</i> complement clauses (NtoCC)	Clausal	Post	means <i>to acquire homes</i>
Dependent phrases	Attributive adjectives (AttrAdj)	Phrasal	Pre	<i>unique</i> contributions
	Pre-modifying nouns (PreModn)	Phrasal	Pre	<i>labour</i> market
	Prepositional phrases (of) (PP_of)	Phrasal	Post	the role <i>of education</i>

Prepositional phrases (others) (PP_others)	Phrasal	Post	variation across social groups
Appositive noun phrases (Apposnp)	Phrasal	Post	Points of Interest (POI)

Result and Discussion

Rhetorical move structures in sociology abstracts across corpora

The overall distribution of rhetorical moves/steps in sociology RA abstracts written by L1 English and L1 Chinese writers is presented in [Table 4](#). Overall, both groups showed similar patterns in rhetorical move structure, with comparable distributions across the five rhetorical moves. In both corpora, M4 (Results) was nearly obligatory, occurring in 97.5% of the abstracts written by L1 English writers and in all abstracts written by L1 Chinese writers, reflecting the importance of reporting research findings in this genre. M2 (Topic) and M5 (Conclusion) also showed high prevalence, each appearing in over 90% of abstracts. At the step level, M1_S2 (emphasizing topic importance), M5_S1 (interpreting results), and M5_S2 (drawing conclusions from results) showed similar distributions, with only minor differences between the two corpora.

Table 4. Distribution of moves / steps in two corpora (n/%)

Move/Step	EAAC	CAAC
M1	145 (90.6)	117 (73.1)
M1_S1	105 (65.6)	88 (55.0)
M1_S2	45 (28.1)	43 (26.8)
M1_S3	81 (50.6)	39 (24.3)
M2	146 (91.2)	141 (88.1)
M3	145 (90.6)	128 (80.0)
M4	156 (97.5)	160 (100.0)
M5	149 (93.1)	141 (88.1)
M5_S1	67 (41.9)	56 (35.0)
M5_S2	48 (30.0)	49 (30.6)
M5_S3	81 (50.6)	97 (60.6)

Notes: “n” is the number of abstracts where a move/step occurs and “%” is the percentage of those abstracts in this corpus

Notably, M3 (Method) also showed a high occurrence rate in both corpora. Rather than functioning as an independent rhetorical unit, it was occasionally integrated with M2 (Topic) or M4 (Results), suggesting a tendency toward rhetorical economy in abstract construction. In the L1 English corpus, M3 was mainly embedded within M2 (Example 1), linking methodological procedures to the research topic. In contrast, in the L1 Chinese corpus, M3 was also observed in combination with M4 (Example 2), presenting methodological procedures alongside research findings. Such integration occurred more frequently in the L1 Chinese corpus, consistent with the findings reported by Li et al. (2025).

- 1) We analyze 43 interviews with Black women aged 18-55 in rural South Africa to explore how the
- 2) An analysis of the event history of 27 provinces from 2000 to 2021 reveals that provinces

Despite these overall similarities, variations were observed in the distribution of M1 (Background) across the

two corpora. L1 English writers showed higher occurrence rates, particularly in M1_S1 (background information) and M1_S3 (gap indication). The difference was more pronounced in M1_S3, which occurred in 50.6% of the abstracts written by L1 English writers but only 24.3% of those produced by L1 Chinese writers. This pattern is consistent with [Lorés-Sanz’s \(2014\)](#) observation that sociology abstracts by English writers tend to provide more explicit contextualization and research gap indication.

This contrast reflects different patterns in the realization of background-related rhetorical functions across the two corpora. The higher frequency of M1_S1 and M1_S3 in the L1 English corpus may indicate a greater tendency to contextualize research and explicitly establish a research niche, consistent with [Swales’ \(1990, 2004\)](#) CARS model. In contrast, previous studies have suggested that L1 Chinese writers of English academic texts may place greater emphasis on establishing research significance rather than identifying gaps ([Loi, 2010](#)). Within the limited space of abstracts, these patterns may represent different approaches to presenting research significance ([Hyland, 1999](#)).

Regarding M5 (Conclusion), L1 English writers showed a higher frequency of M5_S1 (interpreting the results) (41.9%) than L1 Chinese writers (35.0%), while both groups demonstrated similar distributions for M5_S2 (drawing conclusions from results). In contrast, M5_S3 (stating contribution or significance of findings) occurred more frequently in abstracts written by L1 Chinese writers (60.6%) than in those written by L1 English writers (50.6%). This pattern suggests that L1 Chinese abstracts contained more explicit statements of research contribution and significance in the Conclusion move. Such statements may serve to foreground the value and implications of findings. Together with the higher occurrence of M1_S2 (emphasizing topic importance), this pattern is consistent with a more explicit presentation of research significance in both the Introduction and Conclusion moves of L1 Chinese abstracts.

Noun phrasal complexity embedded in rhetorical moves across corpora

Overall, finite dependent clauses occurred more frequently in the EAAC corpus than in the CAAC corpus across most rhetorical moves (see [Figure 1](#)), whereas non-finite structures exhibited more fluctuating patterns in both datasets (see [Figure 2](#)). Dependent phrases showed similar distributional tendencies across rhetorical moves despite variations at the subcategory level (see [Figure 3](#)).

Regarding finite dependent clauses, as shown in [Figure 4](#), finite relative clauses occurred more frequently in the EAAC corpus than in the CAAC corpus across all rhetorical moves. Within the EAAC corpus, these clauses were frequent across most moves, peaking in M5_S1 (interpreting results), followed by M2 (Topic) and M4 (Results), and occurring least frequently in M3 (Method). In contrast, the CAAC showed much lower and more variable frequencies, with the highest occurrence in M1_S3 (gap indication). These distributional differences were supported by Mann-Whitney U tests (see Appendix A), which revealed significant differences in M2 ($U=10770.5$, $Z=-3.731$, $p<.001$), M3 ($U=11442.5$, $Z=-3.664$, $p<.001$), and M4 ($U=8673.5$, $Z=-6.364$, $p<.001$).

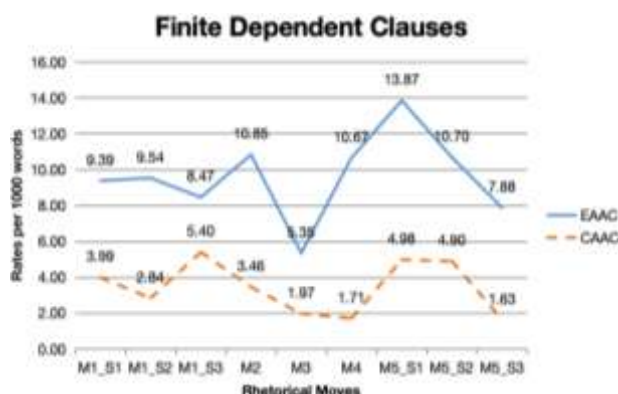


Figure 1. Distribution of finite dependent clauses across rhetorical moves

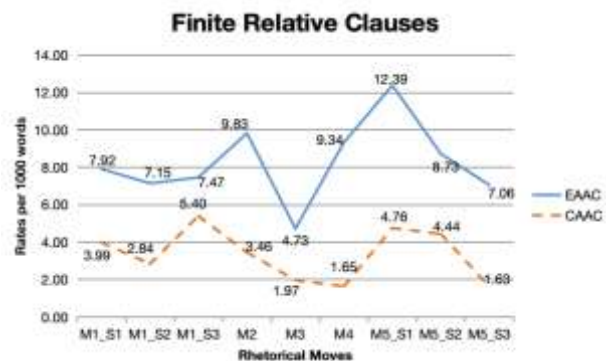


Figure 4. Distribution of finite relative clauses across rhetorical moves

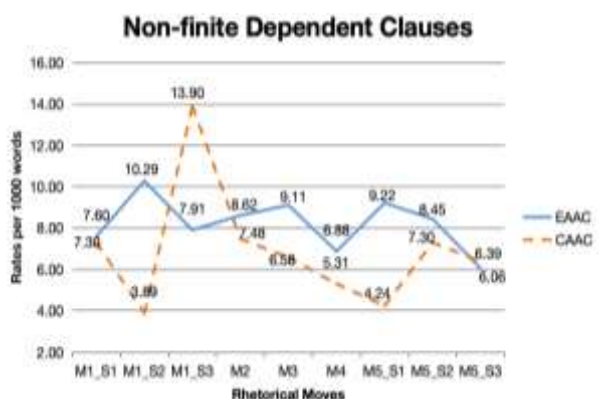


Figure 2. Distribution of non-finite dependent clauses across rhetorical moves

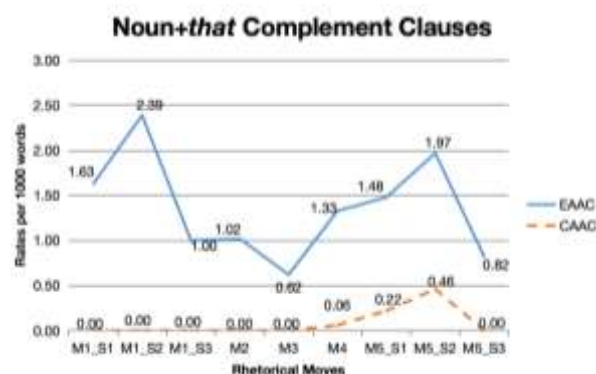


Figure 5. Distribution of noun + that complement clauses across rhetorical moves

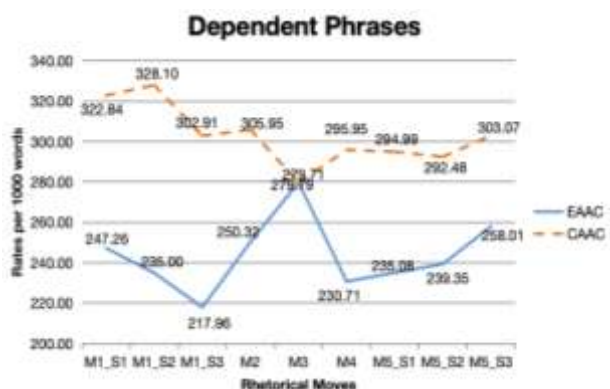


Figure 3. Distribution of dependent phrases across rhetorical moves

These results suggest that finite relative clauses were more strongly associated with several rhetorically important moves in the EAAC corpus. For instance, in M5_S1 of Example 3, L1 English writers utilized the relative clause that harm other marginalized people to specify the consequences of social processes, thereby facilitating interpretation of findings. In contrast, L1 Chinese writers used such structures less frequently, with a slight concentration in M1_S3 (gap indication). As shown in Example 4, the clause that determine compensations served to elaborate the mechanics underlying the identified research gap.

3) Finally, I explain cooptive inclusion—the process of White, male-dominated industries relying on marginalized people to enact policies [that harm other

marginalized people]. (EAAC, M5_S1)

4) They are lack of sufficient attention to the micro-social processes [that determine compensations], especially the tension between the different kinds of losses and the quantitative outcomes of resolution. (CAAC, M1_S3)

Noun+that complement clauses showed a limited distribution in both corpora (see Figure 5).

No statistically significant differences were found between the two groups, suggesting that this structure was not a major source of observed variation in noun phrasal complexity. When used, noun + that complement clauses mainly packaged propositional information into compact noun phrases, particularly through reporting nouns such as evidence, fact, and assumption.

With respect to subcategories of non-finite dependent clauses, the distribution of -ing and -ed non-finite relative clauses largely overlapped across the two corpora (see Figure 6), with no statistically significant differences identified across rhetorical moves/steps. Both types occurred frequently in M3 (Method), where they mainly provided additional descriptive information about research procedures, samples, and related entities while maintaining concise noun phrase structures (Biber et al., 2011). Noun+to complement clauses showed similar distributional patterns across the two corpora, with higher frequencies in M1 (Background) and M5 (Conclusion) (see Figure 7). Although some differences were found in the peak steps between the two corpora, Mann-Whitney U tests revealed no statistically significant differences across rhetorical moves/steps.

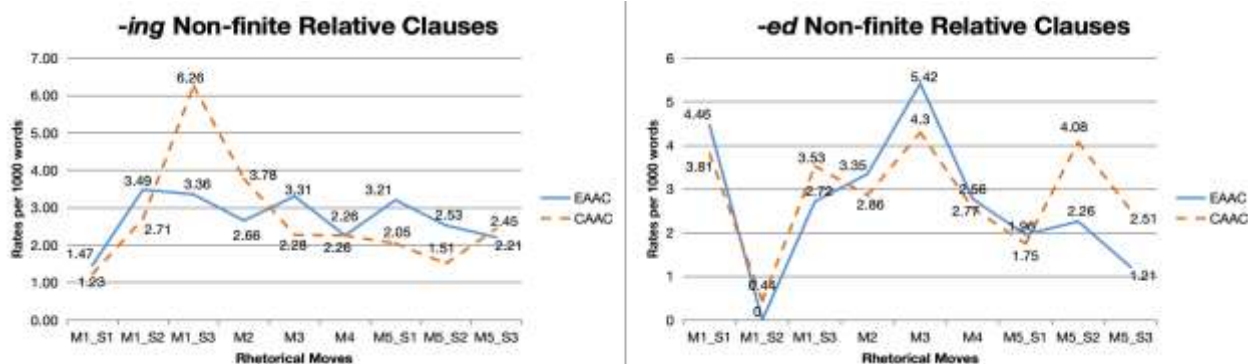


Figure 6. Distribution of non-finite -ed, -ing relatives across rhetorical moves

These structures primarily express purposes, intentions, or goal-directed actions through compact noun phrases. This pattern is consistent with previous observations that academic writing favors phrasal elaboration and dense information packaging over clausal elaboration (Biber & Gray, 2011). As illustrated in Examples 5 and 6, expressions such as technique to avoid system failures and efforts to track the long-term trend condense research objectives into noun phrases.

- 5) Although modularization seemed like a reasonable technique [to avoid system failures], the platforms became vulnerable to persistent gaming. (EAAC, M1_S2)
- 6) Most of the existing studies focus on exploring causal relationships and heterogeneous effects in individual life courses, and efforts [to track the long-term trend] at the macro-level remain scarce. (CAAC, M1_S3)

$p < .001$), M3 ($U = 9778.0$, $Z = -3.731$, $p < .001$), and M4 ($U = 16535.5$, $Z = 4.514$, $p < .001$), although the effect sizes were small. These findings suggest that L1 Chinese writers may rely more on attributive modification in moves involving gap indication, methodological description, and result presentation. As illustrated in Examples 7 and 8, such adjectives specify analytical properties and disciplinary categories, allowing writers to package complex concepts into concise nominal structures.

- 7) We analyze [novel] network data on 16,847 [illegal] drug exchanges between 7205 actors on one [online] [illegal] drug market. (EAAC, M3)

- 8) Based on three [theoretical] perspectives on [private] [economic] development. (CAAC, M3)

Pre-modifying nouns showed similar distributional patterns across the two corpora (see Figure 9), with relatively higher frequencies in M2 (Topic) and M3 (Method).

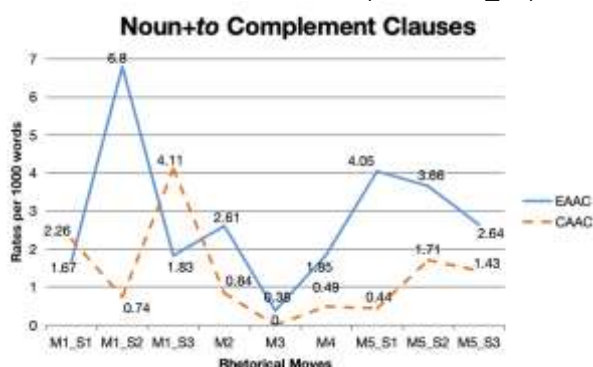


Figure 7. Distribution of noun + to complement clauses across rhetorical move

Among the subcategories of dependent phrases, attributive adjectives were the most frequently used modifier across all rhetorical moves in both corpora (see Figure 8). Their prevalence reflects their function in specifying and classifying nominal concepts while maintaining the compact noun phrase structures (Biber et al., 2011; Biber & Gray, 2016). In sociology research, they are particularly useful for specifying abstract concepts and complex social phenomena.

Both corpora showed a similar distributional pattern, with high frequencies in M1 (Background), a decline in M3 (Method), and a subsequent increase in M4 (Results) and M5 (Conclusion). This pattern suggests that attributive adjectives were more common in moves involving conceptual establishment and interpretation than in methodological descriptions.

Despite this shared pattern, the CAAC corpus showed significantly higher frequencies of attributive adjectives than the EAAC corpus in M1_S3 ($U = 10253.0$, $Z = -3.594$,

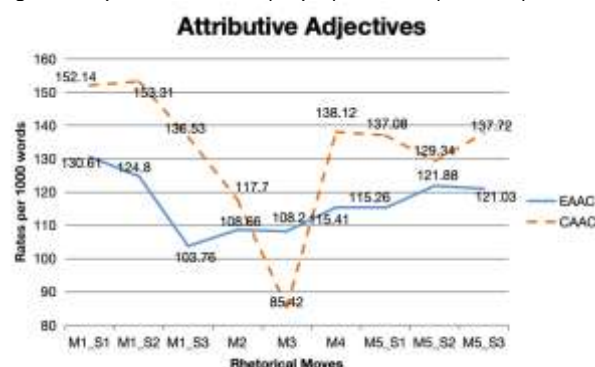


Figure 8. Distribution of attributive adjectives across rhetorical moves

Despite these similarities, the CAAC corpus exhibited significantly higher frequencies than the EAAC corpus in M4 ($U = 17818.0$, $Z = 6.067$, $p < .001$) and M5_S3 ($U = 15539.0$, $Z = 3.746$, $p < .001$), although the effect size was medium only in M4.

The higher frequency of pre-modifying nouns in the CAAC corpus, particularly in M4 (Results), may suggest a greater tendency to use compact nominal expressions when presenting empirical findings. As the Results move often requires repeated reference to variables, participant groups, and analytical categories, noun premodification provides an economical way to express these concepts while maintaining terminological consistency. As illustrated in Examples 9 and 10, such structures package recurring analytical constructs into concise nominal phrases, facilitating efficient reference in the presentation of findings.

- 9) I find that skilled Indian migrants experience occupational upgrades when re-entering their [home] [labor] market, buoyed by foreign human capital and

compensating for US legal [status] restrictions. While immigrants often face initial occupational downgrades and [job] [mobility] constraints in foreign [labor] markets, [return] migrants experience stability, promotions and accelerated rates of [job] mobility at home. (EAAC, M4)

- 10) The findings are as follows: (1) [University] type affects [employment] confidence... (2) [University] location influences [employment] confidence... (3) ...and foreign [language] abilities), the lack of noncognitive skills (such as social skills and [stress] resilience) among ... their [employment] confidence. (4) ...first-tier [city] universities possess... (CAAC, M4)

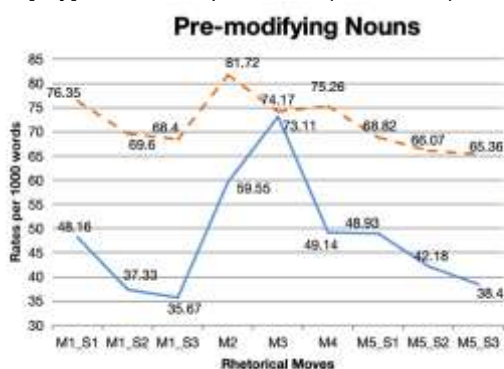


Figure 9. Distribution of pre-modifying nouns across rhetorical moves

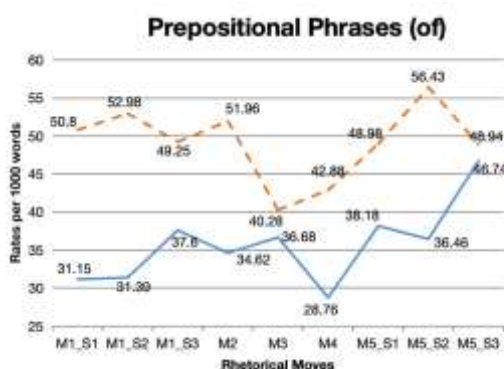


Figure 10. Distribution of prepositional phrases (of) across rhetorical moves

Prepositional phrases, particularly those headed by *of*, constitute an important post-modification resource in academic prose and facilitate information packaging (Biber et al., 1999). As shown in Figure 10, the CAAC corpus displayed higher frequencies of *of*-phrases than the EAAC corpus, with a statistically significant difference found in M4 ($U = 17105.5$, $Z = 5.214$, $p < .001$). This difference reached a medium effect size, indicating that the contrast between the two groups was particularly evident in the Results move.

The higher frequency of *of*-phrases in the CAAC corpus may be associated with greater reliance on relational post-modification in presenting empirical information. In sociology abstracts, *of*-phrases are useful for expressing semantic relationships, including association, attribution, and part-whole relations, within noun phrases. Examples 11 and 12 illustrate how these structures specify such relationships in the two corpora.

- 11) We find that areas... immediate onset [of violence]. However, our analysis [of participation] in the genocide—which uses more accurate measures [of participation] and [of radio coverage]... subnational

levels [of participation]. (EAAC, M4)

- 12) First, significant differences exist among different provinces [of origin]. Second, an increase in... relative mobility levels [of provinces [of origin]] and.... Third, this promoting effect... lower levels [of objective inequality] and a more equal distribution [of educational expectations]. (CAAC, M4)

Regarding prepositional phrases (other than *of*), the two corpora demonstrated similar distributional patterns (see Figure 11). Both groups exhibited higher frequencies of these structures in M3 (Method), and Mann-Whitney U tests revealed no statistically significant differences across rhetorical moves and steps.

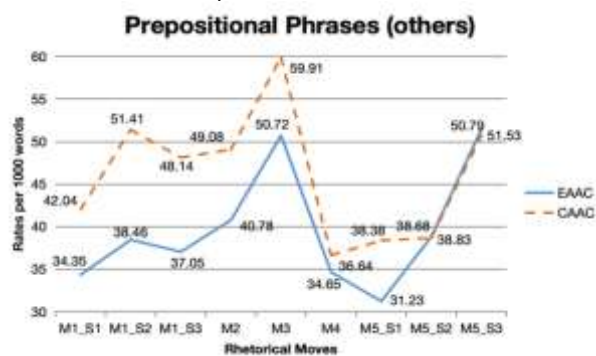


Figure 11. Distribution of prepositional phrases (others) across rhetorical moves

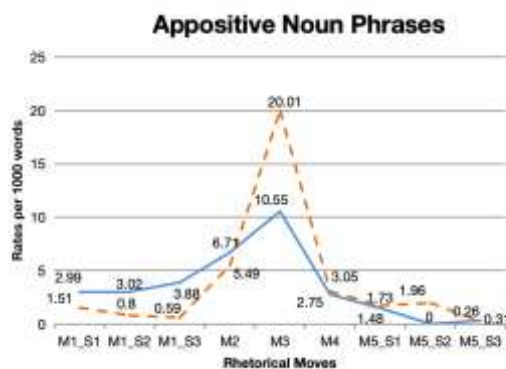


Figure 12. Distribution of appositive noun phrases across rhetorical moves

In this move, from-phrases were frequently used to specify methodological information, including data sources, research populations, and sampling contexts, aligning with the communicative purpose of the Method move (Biber et al., 1999). Examples 13 and 14 illustrate this usage.

- 13) We tested this hypothesis using data from [a survey of state legislators] from [all 50 states ($n = 941$)]. (EAAC, M3)

- 14) Drawing on data from [the 2014 China Family Panel Studies (CFPS)]. (CAAC, M3)

The distribution of appositive noun phrases showed broadly comparable patterns across the two corpora (see Figure 12). In both datasets, these structures occurred most frequently in M3 (Method). Nevertheless, appositive noun phrases showed a higher frequency in CAAC, nearly twice that in EAAC, in this move.

This pattern may suggest that L1 Chinese writers tended to employ appositive noun phrases more frequently in methodological descriptions, particularly for introducing and clarifying discipline-specific terms. However, Mann-Whitney U tests revealed no statistically significant differences between the two groups across rhetorical moves.

Appositive noun phrases served as a compact post-modification resource for providing additional explanatory information without increasing clausal complexity (Biber et al., 1999), particularly in specifying contextual or methodological information. As illustrated in Examples 15 and 16, these structures were used to provide supplementary information about research-related entities, as in the Twin Cities, Minnesota and China Family Panel Studies (CFPS).

15) Drawing on ethnography and interviews with recovering men in the Twin Cities, Minnesota, (EAAC, M3)

16) Using China Family Panel Studies (CFPS) data, (CAAC, M3)

Taken together, the noun phrasal complexity analysis indicates that the most salient differences between the two corpora were concentrated in four features: finite relative clauses, attributive adjectives, noun premodification, and of-prepositional phrases.

L1 English writers showed more frequent use of clausal modification, particularly finite relative clauses, whereas L1 Chinese writers displayed higher frequencies of phrasal modification resources, including attributive adjectives, noun premodifiers, and of-phrases, in several rhetorical moves. By contrast, noun + that complement clauses, non-finite relative clauses, noun + to complement clauses, non-of prepositional phrases, and appositive noun phrases showed broadly comparable distributions across the two corpora, suggesting shared functional requirements in sociology abstract construction.

These findings reveal that noun phrasal complexity is closely connected with the rhetorical functions realized in different moves rather than representing an independent measure of grammatical elaboration. By integrating move analysis with noun phrasal complexity analysis, this study extends previous research that has often treated rhetorical move structures and linguistic complexity separately, demonstrating how grammatical choices serve genre-specific communicative purposes in academic writing.

Conclusion

This study examined rhetorical move structures and noun phrasal complexity in sociology RA abstracts written by L1 Chinese and L1 English writers by integrating a move-based framework with a fine-grained taxonomy of noun phrasal complexity. The findings show that, although both groups generally followed the conventional rhetorical move structure of sociology abstracts, they differed in their realization of information packaging across move contexts. L1 Chinese writers relied more on phrasal modification resources, whereas L1 English writers showed greater use of clausal modification resources, particularly in result- and conclusion-oriented moves. These patterns suggest that rhetorical choices may be associated with differences in

writers' linguistic and academic backgrounds while also reflecting shared disciplinary conventions. Therefore, linguistic complexity should be examined in relation to genre-specific rhetorical functions rather than as a decontextualized measure of structural elaboration.

The study offers implications for EAP and ESP pedagogy. Specifically, the findings highlight the need to align noun phrasal complexity instruction with rhetorical move structures rather than treating grammatical complexity as a collection of decontextualized forms. Because different rhetorical moves involve different communicative demands, academic writing instruction should address not only how complex structures are formed but also how they function within particular rhetorical contexts. Disciplinary exemplars can help learners recognize how experienced writers deploy noun modification resources to organize disciplinary information and fulfill genre-specific communicative purposes.

Several limitations of the present study should nevertheless be acknowledged. First, the investigation focused exclusively on sociology RA abstracts within a relatively limited corpus and publication period. Consequently, the findings may not be fully generalizable to other disciplinary contexts or broader academic writing practices. Second, although the annotation process demonstrated high intra-coder consistency, independent inter-coder reliability testing was not conducted. Third, while the corpus-based comparison identified associations between writer groups and linguistic-rhetorical patterns, it cannot determine whether these differences are attributable to writers' linguistic backgrounds alone, as other factors such as academic training, publication practices, and journal conventions may also contribute. Future research could address these limitations by examining larger and more diverse corpora, incorporating more detailed author-background variables, and involving multiple trained coders to enhance the generalizability and reproducibility of findings.

Author contributions

Qiaoqiao Chen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing-original draft, Writing-review & editing. Zhijun Li: Conceptualization, Data curation, Investigation, Methodology, Validation, Writing-original draft, Writing-review & editing.

Acknowledgements

We would like to express our thanks to the editor and anonymous reviewers for their valuable comments on this paper.

References

- 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>
- Arizavi, S., Namdari, N., & Mousavi, S. A. (2015). Nominalization in the research article discussion sections of local and international journals of applied linguistics. *Iranian Journal of Applied Linguistics*, 18(1), 1–28. <https://doi.org/10.18869/acadpub.ijal.18.1.1>
- Becher, T., & Trowler, P. R. (2001). *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines* (2nd ed.). SRHE/Open University Press.
- Bhatia, V. K. (1993). *Analyzing Genre: Language Use in Professional Settings*. Longman.
- Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics*, 15(2), 223–250. <https://doi.org/10.1017/s1360674311000025>
- Biber, D., & Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511920776>

- 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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.
- Bondi, M. (2014). Changing voices: Authorial voice in abstracts. In M. Bondi & R. Lorés Sanz (Eds.), *Abstracts in Academic Discourse: Variation and Change* (pp. 243–269). Peter Lang. <https://doi.org/10.3726/978-3-0351-0701-2/23>
- Busch-Lauer, I. A. (1995). Abstracts in German medical journals: A linguistic analysis. *Information Processing and Management*, 31(5), 769–776. [https://doi.org/10.1016/0306-4573\(95\)00024-b](https://doi.org/10.1016/0306-4573(95)00024-b)
- 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, 100996. <https://doi.org/10.1016/j.jeap.2021.100996>
- Cortés, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12(1), 33–43. <https://doi.org/10.1016/j.jeap.2012.11.002>
- Cumming, A. (2013). Multiple dimensions of academic language and literacy development. *Language Learning*, 63, 130–152. <https://doi.org/10.1111/j.1467-9922.2012.00741.x>
- Darabad, A. M. (2016). Move analysis of research article abstracts: A cross-disciplinary study. *International Journal of Linguistics*, 8(2), 125–140. <https://doi.org/10.5296/ijl.v8i2.9379>
- Gong, H., Le, T. N. P., & Buckingham, L. (2025). Lexical bundles across IMRD-structured Medicine research article sections: A within-register perspective. *Journal of English for Academic Purposes*, 74, 101487. <https://doi.org/10.1016/j.jeap.2025.101487>
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2013). *Halliday's Introduction to Functional Grammar* (4th ed.). Routledge. <https://doi.org/10.4324/9780203431269>
- Hu, Y., & He, Q. (2023). A corpus-based study of the distributions of adnominals across registers and disciplines. *Journal of Quantitative Linguistics*, 30(2), 183–203. <https://doi.org/10.1080/09296174.2023.2209487>
- Huckin, T. N. (2001). Abstracting from abstracts. In M. Hewings (Ed.), *Academic Writing in Contexts: Implications and Applications* (pp. 93–103). Bloomsbury Academic. <https://doi.org/10.5040/9781474211734>
- Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. In C. N. Candlin & K. Hyland (Eds.), *Writing: Texts, Processes and Practices* (pp. 99–121). Routledge. <https://doi.org/10.4324/9781315840390-6>
- Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. Longman.
- Hyland, K. (2004). *Genre and Second Language Writing*. University of Michigan Press. <https://doi.org/10.3998/mpub.23927>
- Kobayashi, H., & Rinnert, C. (2008). Task response and text construction across L1 and L2 writing. *Journal of Second Language Writing*, 17(1), 7–29. <https://doi.org/10.1016/j.jslw.2007.08.004>
- 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>
- Lam, Y. S., & Tan, H. (2012). A comparative study of the rhetorical moves in abstracts of published research articles and students' term papers in the field of computer and communication systems engineering. *International Journal of Applied Linguistics and English Literature*, 1(7), 40–50. <https://doi.org/10.7575/ijael.v.1n.7p.40>
- Lan, G., Lucas, K., & Sun, Y. (2019). Does L2 writing proficiency influence noun phrase complexity? A case analysis of argumentative essays written by Chinese students in a first-year composition course. *System*, 85, 102116. <https://doi.org/10.1016/j.system.2019.102116>
- Lan, G., Zhang, Q., Lucas, K., Sun, Y., & Gao, J. (2022). A corpus-based investigation on noun phrase complexity in L1 and L2 English writing. *English for Specific Purposes*, 67, 4–17. <https://doi.org/10.1016/j.esp.2022.02.002>
- Leki, I., Cumming, A., & Silva, T. (2008). *A Synthesis of Research on Second Language Writing in English*. Routledge.
- Li, L. J., & Ge, G. C. (2009). Genre analysis: Structural and linguistic evolution of the English-medium medical research article (1985–2004). *English for Specific Purposes*, 28(2), 93–104. <https://doi.org/10.1016/j.esp.2008.12.004>
- Li, Z. J., Lin, J. Y., & Xu, J. F. (2025). Promoting research: Academic hypotheses embedded in the rhetorical move structure of sociology research article abstracts. *Journal of English for Academic Purposes*, 76, 101535. <https://doi.org/10.1016/j.jeap.2025.101535>
- Loi, C. K. (2010). Research article introductions in Chinese and English: A comparative genre-based study. *Journal of English for Academic Purposes*, 9(4), 267–279. <https://doi.org/10.1016/j.jeap.2010.09.004>
- Lorés-Sanz, R. (2014). Lost (and gained) in translation: A contrastive (English/Spanish) analysis of rhetorical and lexicogrammatical patterns in sociology research article abstracts. In M. Bondi & R. Lorés Sanz (Eds.), *Abstracts in Academic Discourse: Variation and Change* (pp. 85–109). Peter Lang. <https://doi.org/10.3726/978-3-0351-0701-2/14>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36–62. <https://doi.org/10.5054/tq.2011.240859>
- Lu, X., Casal, J. E., & Liu, Y. (2020). The rhetorical functions of syntactically complex sentences in social science research article introductions. *Journal of English for Academic Purposes*, 44, 100832. <https://doi.org/10.1016/j.jeap.2019.100832>
- Maswana, S., Kanamaru, T., & Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand*, 2, 1–11. <https://doi.org/10.1016/j.amper.2014.12.002>
- Pan, F., & Yang, Y. (2024). Diachronic changes in the phrasal complexity of research articles (1970–2020): A cross-disciplinary investigation. *Scientometrics*, 129(7), 4395–4421. <https://doi.org/10.1007/s11192-024-05060-w>
- Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for academic purposes students. *Journal of English for Academic Purposes*, 14, 48–59. <https://doi.org/10.1016/j.jeap.2013.12.001>
- Parviz, M., & Lan, G. (2023). A corpus-based investigation of phrasal complexity features and rhetorical functions in data commentary. *Journal of Language and Education*, 9(3), 90–109. <https://doi.org/10.17323/jle.2023.16044>
- 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, 149–183. <https://doi.org/10.35869/vial.v0i20.4357>
- Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, 10(2), 231–250. <https://doi.org/10.1177/1461445607087010>
- Ruan, Z. (2018). Structural compression in academic writing: An English-Chinese comparison study of complex noun phrases in research article abstracts. *Journal of English for Academic Purposes*, 36, 37–47. <https://doi.org/10.1016/j.jeap.2018.09.001>
- Santos, M. B. D. (1996). The textual organization of research paper abstracts in applied linguistics. *Text-Interdisciplinary Journal for the Study of Discourse*, 16(4), 481–500. <https://doi.org/10.1515/text.1.1996.16.4.481>
- Shao, Z. M., Zhang, H. B., Zhang, J. Q., & Tsai, K. J. (2022). Phrasal complexity in English argumentative writing: Variations across Chinese STEM versus English majors' production and EFL textbook essays. *Journal of English for Academic Purposes*, 55, 101070. <https://doi.org/10.1016/j.jeap.2021.101070>
- Shen, C., Guo, J., Shi, P., & Liu, H. (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), e0292688. <https://doi.org/10.1371/journal.pone.0292688>
- Sükan, S., & Mohammadzadeh, B. (2022). Challenges of writing theses and dissertations in an EFL context: Genre and move analysis of abstracts written by Turkish MA and Ph.D. students. *Frontiers in Psychology*, 13, 925420. <https://doi.org/10.3389/fpsyg.2022.925420>
- Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Swales, J. M. (2004). *Research Genres: Explorations and Applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524827>
- Swales, J. M., & Feak, C. B. (2009). *Abstracts and the Writing of Abstracts*. The University of Michigan Press. <https://doi.org/10.3998/mpub.309332>
- Tankó, G. (2017). Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. *Journal of English for Academic Purposes*, 27, 42–55. <https://doi.org/10.1016/j.jeap.2017.04.003>
- Terblanche, L. (2009). A comparative study of nominalisation in L1 and L2 writing and speech. *Southern African Linguistics and Applied Language Studies*, 27(1), 39–52. <https://doi.org/10.2989/salals.2009.27.1.4.752>
- Thongyoi, K., & Poonpon, K. (2020). Phrasal complexity measures as predictors of EFL university students' English academic writing proficiency. *REFlections*, 27(1), 44–61. <https://doi.org/10.61508/refl.v27i1.241750>
- Tseng, F. P. (2011). Analysis of move structure and verb tense in research article abstracts in applied linguistics. *International Journal of English Linguistics*, 1(2), 27–39. <https://doi.org/10.5539/ijel.v1n2p27>
- Wang, Y., Ge, T., & Ren, Z. (2022). Noun phrasal complexity in computer science conference abstracts: A corpus-based study. *International Journal of Computer-Assisted Language Learning and Teaching*, 12(1), 1–17. <https://doi.org/10.4018/ijcall.311096>
- Wang, Y. X., & Jiang, J. Y. (2024). Exploring the development of noun phrase complexity in L2 English writings across two genres. *Assessing Writing*, 62, 100892. <https://doi.org/10.1016/j.asw.2024.100892>
- Xie, S., & Mi, C. (2023). Promotion and caution in research article abstracts: The use of positive, negative and hedge words across disciplines and rankings. *Learned Publishing*, 36(2), 249–265. <https://doi.org/10.1002/leap.1515>