International Journal of Applied Linguistics

WILEY

ORIGINAL ARTICLE OPEN ACCESS

More Complex Constrained Language? An Investigation on Nominal Complexity in Constrained Spoken English Varieties of ESL, EFL, and Interpreted English

Jiaxin Chen[1] | Yao Yao[2,3] | Dechao Li[3]

[1]Department of Foreign Languages, China Agricultural University, Beijing, PR China | [2]School of Foreign Studies, Xi'an Jiaotong University, Shaanxi, PR China | [3]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

Correspondence: Dechao Li (dechao.li@polyu.edu.hk)

Funding: The research was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 15602621).

Keywords: Constrained English varieties | interpreted English | L2 spoken English | language proficiency | nominal complexity | task topic | 名词复杂度 | 受限英语变体 | 语言水平 | 任务主题 | 口译英语

ABSTRACT

While previous research has established nominal complexity as a key indicator of grammatical complexity, its multi-dimensional manifestation in constrained varieties remains underexplored. This study investigates the distribution of complex nominal structures in constrained spoken English varieties, specifically English as a Second Language (ESL), English as a Foreign Language (EFL), and Interpreted English, compared to non-constrained, L1 English. Drawing on data from the International Corpus Network of Asian Learners of English (ICNALE) and a self-compiled corpus of interpreted English, this study examines various structural types of noun modification in constrained English varieties and explores how language proficiency and task topic influence patterns of nominal complexity. Results indicate a general tendency across constrained varieties to favor phrasal over clausal noun modification, suggesting a potential impact of limited exposure to diverse registers and modalities of English. Interestingly, while language proficiency level does not significantly affect the distribution of complex nominal structures, task topic emerges as a significant factor influencing their deployment, highlighting the role of communicative contexts in shaping grammatical choices. By shedding light on language production within the constraints of bilingual activation, this study contributes valuable insights to theories of second language acquisition and informs pedagogical approaches in ESL/EFL and interpreter training.

摘要

尽管已有研究证实名词复杂度是衡量语法复杂性的核心指标, 但其在受限语言变体中的多维度表现仍待深入探索。本研究以非受限的母语英语为参照, 聚焦作为第二语言的英语 (English as a Second Language, ESL) 、作为外语的英语 (English as a Foreign language, EFL) 及口译英语 (Interpreted English) 三类受限英语口语变体, 探讨复杂名词结构的分布特征及其影响因素。基于"亚洲英语学习者国际语料库网络" (ICNALE) 和自建口译英语语料库, 本研究系统考察了受限英语变体中名词修饰的多种结构类型, 并探究语言水平与任务主题如何影响名词复杂度的使用模式。研究发现, 受限英语变体普遍倾向于使用短语修饰而非从句修饰, 这一现象可能反映了学习者接触英语不同语域和形式的有限性。此外, 尽管语言水平对复杂名词结构的分布无显著影响, 但任务主题是影响复杂名词结构使用的重要因素, 凸显了交际语境对语法选择的关键作用。本研究通过揭示双语激活制约下的语言产出模式, 为二语习得理论提供了新视角, 并为ESL /EFL教学及口译培训提供了实践参考。

1 | Introduction

The study of constrained language, or constrained communication, represents a burgeoning field of inquiry that bridges disciplines such as language contact, language varieties, second language (L2) acquisition, and translation studies (Chesterman 2004; Granger 2015; Ulrych and Murphy 2008). This line of research focuses primarily on the language production of bilingual or multilingual individuals, recognizing that their linguistic choices are influenced by the cognitive and socio-cultural demands of juggling multiple languages (Kotze 2022; Lanstyák and Heltai 2012). Constrained language is often characterized by features such as simplified lexical-grammatical encodings and a preference for conventional linguistic choices, attributed to the higher cognitive demand in the bilingual brain when managing multilingual resources. However, empirical evidence supporting these features remains mixed, underlining the complex interplay of cognitive and social factors in shaping language production (De Sutter and Lefer 2020; Kajzer-Wietrzny and Ivaska 2020; Kruger and Van Rooy 2016; Liu et al. 2023; Chen et al. 2024).

Despite the growing interest in constrained language, one area that remains underexplored is the complexity of nominal structures. English noun phrases, in particular, are known for their potential for intricate grammatical structures, incorporating a range of pre- and post-modifications (Quirk et al. 1985). Research in L2 acquisition and regional variation in English has shown that the complexity of noun phrases not only signifies grammatical sophistication but also serves as a marker differentiating production modalities (spoken vs. written) and language development stages (Biber and Gray 2010; Schilk and Schaub 2016; Biber et al. 2011). Notably, nominal complexity remains underexplored in research on constrained language, despite evidence suggesting its potential significance. While Kruger and van Rooy (2016, 2018) revealed important parallels between translated English and non-native indigenized varieties in their increased formality and information explicitation as evidenced by the more frequent use of specific linguistic features including nominal structures, systematic investigation of nominal complexity is lacking. This gap is particularly noteworthy given Ivaska et al.'s (2022, 152) explicit call to investigate "the tendency for phrasal versus clausal elaboration… focusing on nouns and verbs, and structures around them."

The present study addresses this gap by examining nominal complexity within the context of constrained language production, particularly focusing on spoken learner language across constrained varieties of English. Rather than treating grammatical complexity as a unitary construct, we adopt a comprehensive analytical framework that captures the intricate nature of nominal structures by accounting for their multi-dimensionality (Biber et al. 2023). This investigation aims to identify the distributional patterns of nominal complexity across these varieties, exploring the effect of both shared and unique cognitive and social constraints involved. Ultimately, this study seeks to contribute valuable insights to the theoretical understanding of constrained language while also informing pedagogical approaches in the second language acquisition and interpreter training context.

2 | Literature Review

2.1 | Interpreted English and L2 Englishes as Constrained Varieties of English

Constrained communication refers to "the language produced in communicative contexts characterized by particularly conspicuous constraints" (Kruger and Van Rooy 2016, 27). Empirical studies have identified commonalities in languages influenced by bilingualism, where language users are subject to the impacts of two languages. Kotze (2022) expanded this concept by outlining five dimensions that condition constrained language use, including register and modality, production mode, and languages activated during production, as well as the task expertise and language proficiency of the participants involved.

Despite the acknowledged influence of production mode (spoken vs. written) on language use (Biber 1988), research on constrained language has predominantly focused on written language, leaving spoken forms relatively underexplored. Notable work by Kajzer-Wietrzny (2018) revealed that both interpreters and L2 English speakers use the optional complementizer more frequently than L1 speakers. Building on this, Kajzer-Wietrzny (2022) further conducted intermodal comparisons of cohesive devices in constrained and non-constrained English across spoken and written modes. The findings indicate that both L2 and mediated productions diverge from L1 production, albeit in different ways. For instance, interpreters and translators tend to use similar transitional phrases, whereas L2 speakers favor enumerative devices in spoken discourse. Additionally, Kajzer-Wietrzny and Ivaska (2020) observed that interpreted languages exhibit lower lexical density compared to translated and L2 speech, suggesting lexical simplification and a preference for explicit grammatical encoding in constrained spoken varieties. More simplified syntactic structures were also identified in constrained L2 speaking and interpreted English, as shown in Liu et al. (2023). These findings underscore the significant influence of production mode on language use in constrained communication.

A further contribution of this study lies in its emphasis on the multi-dimensionality of grammatical complexity, building on the framework developed by Biber and colleagues (Biber et al. 2020, 2023). Their work has established that grammatical complexity is fundamentally multi-dimensional, where syntactic functions are as crucial as structural types in analyzing complexity patterns—an insight that challenges traditional approaches relying on omnibus measures. We argue that this multi-dimensionality extends beyond written texts to language production in general. By examining nominal complexity and its interaction with various structural types in spoken language, this study moves beyond treating grammatical complexity as a unitary construct, instead investigating different complexity features' functions and patterns across spoken language varieties.

2.2 | Complex Nominal Structures in Learner Production

Phrasal elaboration, particularly nominal elaboration, is a fundamental aspect of syntactic complexity that contributes to the

informational richness and specificity of phrases (Kyle and Crossley 2018). This process involves the expansion or modification of nouns to form complex nominal structures by incorporating various pre-modifiers (e.g., nouns, possessives, and attributive adjectives) and post-modifiers (e.g., prepositional phrases, verb phrases, and relative clauses). Languages characterized by a high frequency of complex nominal structures are recognized for their high information density. A higher frequency of complex nominal structures contributes to increased information density, which can lead to greater cognitive load during language comprehension and production, suggesting a correlation between nominal complexity and overall linguistic complexity.

Research consistently highlights the integral role of complex nominal structures in language development and proficiency (Norris and Ortega 2009; Biber et al. 2011). Norris and Ortega (2009) contend that phrasal complexity serves as an accurate indicator of grammatical complexity, reflecting advanced stages of language use and maturity. This perspective is further supported by Biber et al. (2011), who suggest that while dependent clauses are prevalent in the early stages of writing development, the use of phrasal modifiers emerges as a marker of greater syntactic complexity in advanced learners. Empirical evidence reinforces these findings, demonstrating that higher linguistic proficiency often correlates with an increased use of complex nominal structures in both written and spoken discourse (Ansarifar et al. 2018; Bao 2024). Clausal and phrasal elaboration could distinguish spoken from written language (Biber 1988; Halliday 1990). Spoken registers are characterized by clausal structures for noun modification, while written registers are typified with often nominal phrases. Empirical evidence supports this distinction: Kormos (2014) found EFL learners used more modifiers per noun phrase in written tasks than in speaking tasks. Similarly, Cho (2018) observed lower complex nominal structure per T-unit in speeches compared to writings by Korean university-level EFL learners.

In the context of spoken language specifically, research has shown that higher learner proficiency manifests differently in terms of grammatical complexity. Studies have found correlations between proficiency and complexity primarily at the lexical level (Lu 2012; De Clercq 2015) or through holistic syntactic measures such as mean length of T-unit and clauses per T-unit (Iwashita et al. 2008; Gan 2012) and subordination (Seedhouse et al. 2014). While De Clercq and Housen (2017) expanded their analysis to include mean length of noun phrase alongside length of AS-unit and clauses per AS-unit in examining the relationship between proficiency and complexity, nominal complexity was not found to be a strong discriminating factor between proficiency levels in spoken L2 English and French. Yet the relative scarcity of nominal complexity in spoken language makes it a potentially more meaningful indicator when it does occur, possibly signaling advanced proficiency or specific discourse functions.

While learner proficiency is a well-established driver of nominal complexity, the impact of communicative task topics remains underexplored (Jucker 1992; Yang et al. 2015; Yoon 2017). Research indicates that tasks requiring specificity or technical knowledge often elicit more complex nominals, possibly due to the need for precise information delivery (Jucker 1992). Yang et al. (2015) found that learners use more complex noun phrases when discussing familiar topics, suggesting a link between topic familiarity and syntactic complexity. This may result from the reduced cognitive load in such a setting, which allows more resources for complex language production. Yoon (2017) noted increased nominal complexity when task topics aligned with learners' interests or expertise. These studies represent initial findings, and further research is needed to understand the interplay between task topics and learner proficiency in shaping nominal complexity.

Thus, to address the identified research gaps, this study aims to explore nominal complexity across three constrained spoken varieties of English at the learner level: ESL, EFL, and interpreted English. Recognizing the potential influence of language proficiency and task topic, the study is structured around three key research questions:

1. What are the distributions of complex nominal structures in the three constrained spoken English varieties (i.e., ESL, EFL, and interpreted English) at the learner level in comparison to non-mediated, L1 spoken English?
2. To what extent do learners' language proficiency and task topic affect the use of complex nominal structures across these varieties?
3. What are possible explanations to the similarities and differences across these constrained varieties regarding the patterns of complex nominal structures?

3 | Data and Methodology

3.1 | Data

The current study investigates nominal complexity across three constrained varieties of spoken English: interpreted English (INT), ESL, and EFL, with non-mediated, L1 English serving as a baseline. This investigation specifically hones in on the Chinese-English language pair, focusing on learner productions within these varieties.

While acknowledging that the boundaries between language varieties such as ESL and EFL are not discrete and that viewing languages as bounded entities is problematic from a translanguaging perspective (Wei 2018), we adopt an operational distinction following Gilquin's (2024) framework. Gilquin conceptualizes various English varieties along a continuum between New Englishes and Learner Englishes, characterized by different degrees of interaction with local languages. Within this framework, we classify Mainland China as an EFL context, where English is primarily acquired through formal education with limited exposure to daily life. Hong Kong, in contrast, represents an ESL context, characterized by greater English exposure in daily communication and the language's official status in the region.

To ensure a robust and representative analysis, data were drawn from two primary sources: the International Corpus Network of Asian Learners of English (ICNALE) and in-class interpreting recordings. The ESL, EFL, and L1 English datasets were extracted

TABLE 1 | Descriptive information of non-constrained and constrained corpora.

	L1	ESL		EFL	INT	
Language proficiency	**C2**	**B2**	**B1**	**B2**	**B1**	**C1**
Number of texts	200	104	92	40	156	782
Total word tokens	30,630	12,195	10,952	4,191	15,706	78,884
Mean text length	153.15	117.26	119.04	104.78	100.68	100.87
Text length SD	29.11	20.58	22.09	16.02	24.27	45.59

TABLE 2 | Constrained matrix of data.

	L1	ESL	EFL	INT
Language activation	Monolingual	Bilingual (Cantonese & English)	Bilingual (Mandarin Chinese & English)	Bilingual (Mandarin Chinese & English)
Language proficiency	L1 English speakers	Intermediate-level English speakers	Intermediate-level English speakers	Intermediate-Advanced English speakers
Modality & register	Informal/general speech	Informal/general speech	Informal/general speech	Informal/general speech
Text production	Independent production	Independent production	Independent production	Mediated production
Task expertise	Learner level	Learner level	Learner level	Learner level

from the ICNALE corpus (Ishikawa 2023), which comprises a collection of controlled essays and speeches produced by English learners across various Asian countries and regions. This corpus provides transcripts of the spoken data along with metadata like speakers' first language and English proficiency levels, as assessed by the Common European Framework of Reference (CEFR). Six-second monologues on two topics—"Part-time Job (PTJ)" and "Smoking (SMK)"—produced by learners from Hong Kong (for ESL) and mainland China (for EFL) were selected. To control for language proficiency within each variety, only data from learners at CEFR levels B1 (intermediate, i.e., B1-upper in ICNALE) and B2 (upper-intermediate, i.e., B2+ in ICNALE) were included. Monologues on the same topics by L1 English learners at CEFR level C2 were selected to represent L1 English. It is important to note that the CEFR levels reported in Ishikawa (2023) were determined using a self-developed conversion table, which maps students' external test scores (e.g., TOEFL, IELTS) or vocabulary size test scores to the CEFR scale (Ishikawa 2023, 6).

Data representing the INT variety were collected from in-class interpreting recordings at a master's program in Translation and Interpreting at a Hong Kong university. The source speeches for these interpreting tasks covered general topics such as films, music, and educational issues in Hong Kong. All interpreter trainees possessed an IELTS score of 6.5 or above, equivalent to a CEFR level of C1 or higher.

Prior to analysis, all transcripts underwent a rigorous cleaning process. This process, conducted manually by the first two authors, involved removing disfluency markers (e.g., "uh" and "umm"), correcting truncated words (e.g., "smo-ke"), and rectifying spelling errors to ensure the accuracy of subsequent linguistic analysis. A comprehensive overview of the data is presented in Table 1.

Following Kotze's (2022) framework, a constraint matrix pertinent to our dataset is detailed in Table 2. This matrix highlights bilingual activation as the primary commonality among INT, EFL, and ESL, a factor absent in L1 English production. Crucially, INT differs from other varieties due to its inherent nature as mediated text production. Given our focus on examining the linguistic impacts of bilingual activation, production mode, and language proficiency within these constrained language settings, we maintain consistency across other potential constraint dimensions. This is achieved by drawing on data representing general speech produced by individuals at the learner level across all varieties.

3.2 | Nominal Complexity Measures and Statistical Analysis

Holistic indices, such as complex nominals per T-unit and per clause, have been shown to be effective measures of nominal complexity in evaluating language performance and distinguishing between language proficiency levels (Lu 2010, 2017). While these holistic indices offer valuable insights, they do not differentiate between types of modifications, merging categories like noun phrases and relative clauses, and thus may obscure nuanced patterns in language development and variation that become

TABLE 3 | Nominal complexity measures.

Classification	Label	Measure	Calculation
Phrasal noun modification	amod	An adjective that modifies a noun or noun phrase	Number of adjectival modifiers per nominal
	prep	A prepositional phrase that modifies a noun or noun phrase	Number of prepositions per nominal
	poss	A possessive pronoun or noun with a possessive "s" that modifies a noun or noun phrase	Number of possessives per nominal
	nn	A noun that modifies a noun or noun phrase	Number of nouns as a nominal dependent per nominal
	advmod	An adverb that modifies a noun or noun phrase	Number of (non-clausal) adverbial modifiers per nominal
Clausal noun modification	rcmod	A relative clause is a clause that modifies a noun or noun phrase, and is often marked by a "wh" word	Number of relative clause modifiers per nominal
	vmod	A non-finite verb or verb phrase that modifies a noun or noun phrase	Number of verbal modifiers per nominal
Noun phrase coordination	conj	conjunction "and" and "or" as a nominal dependent per nominal	Number of conjunction "and" and "or" as a nominal dependent per nominal

evident when considering the specific types of modifications employed.

In response to Biber et al.'s (2020) call for more linguistically interpretable measures, we adopted eight fine-grained indices to capture distinct types of phrasal or clausal modification within noun phrases. These indices encompass noun phrases pre- or post-modified by adjectives, possessives, nouns, adverbials, prepositional phrases, non-finite verbal phrases, and relative clauses, as well as coordinated noun phrases conjoined by "and" and "or." Although nominal complexity measures are traditionally applied to written texts, their extension to spoken language—particularly constrained varieties such as interpreted language—is both justified and methodologically appropriate. First, grammatical complexity was increasingly recognized as a multidimensional construct, extending beyond written texts to encompass spoken language production (ibid.; Biber et al. 2023). Second, previous research has established that interpreted languages, situated on a literate-oral continuum, often exhibit features typically associated with written or "literate" registers (Shlesinger and Ordan 2012; Yao et al. 2024). This literate dimension of interpreted languages thus provides a solid rationale for employing measures traditionally reserved for written texts, thereby reinforcing the suitability of nominal complexity metrics within the present study's analytical framework.

Instead of raw counts, we employ the relative frequencies of these structures as the primary metric for assessing nominal complexity. This approach, detailed in Table 3, allows for meaningful comparison across varieties while controlling for potential discrepancies in corpus size. Data parsing and frequency calculations were performed using TAASSC 1.3.8 (Kyle and Crossley 2018). The eight structures can be classified into three categories: phrasal noun modification, clausal noun modification, and noun phrase coordination. This categorization enables a multi-layered approach to nominal complexity in the different varieties of English.

To explore the impact of English variety on nominal complexity (RQ1), the descriptive statistics for each sub-corpus were first calculated using TAASSC results. Next, non-parametric Kruskal—Wallis tests were conducted due to the non-normal distribution of the data. These tests compared the relative frequencies of the eight complex nominal structures across the three constrained varieties (INT, ESL, EFL) and L1 English. Subsequently, post-hoc Dunn's tests with Bonferroni correction were implemented to pinpoint specific significant pairwise differences between the varieties.

To examine the influence of language proficiency and task topic on nominal complexity (RQ2), linear mixed-effects regression models were employed for each nominal structure type. The models included Topic (with two levels: PTJ and SMK) and CEFR (with three levels: B1, B2, and C2) as predictors. To account for the repeated measures from individual participants, Student ID was included as a random effect in the model formulation: Dependent variable ~ Topic + CEFR + (1|Stu_ID). Importantly, to mitigate potential multicollinearity—for example, the exclusive association of interpreted English data with the C1 level and the general topic (GEN), potentially leading to collinearity between the Variety and CEFR/Topic factors—we restricted this part of the analysis to EFL, ESL, and L1 English. This restriction ensures a consistent distribution of topics across varieties. All statistical modeling was conducted in R using the *lme4* package (Bates et al. 2014). The effect size and effect plots were conducted using the *MuMIn* and *effects* packages. Diagnostic checks of our mixed-effects models revealed that only the adjective modification model fully met residual normality assumptions. While our large sample size provides some robustness against non-normality (Hayashi 2000), these departures from distributional

TABLE 4 | Descriptive statistics of the eight nominal complexity measures.

	L1	ESL	EFL	INT
	Median (SD)	Median (SD)	Median (SD)	Median (SD)
amod	0.1277 (0.0606)	0.1727 (0.0790)	0.1571 (0.0876)	0.1349 (0.0967)
prep	0.0862 (0.0550)	0.0990 (0.0573)	0.0909 (0.0703)	0.1333 (0.0826)
poss	0.0465 (0.0421)	0.0588 (0.0471)	0.0667 (0.0556)	0.0667 (0.0674)
nn	0.0458 (0.0390)	0.0542 (0.0539)	0.0351 (0.0584)	0.0476 (0.0898)
advmod	0.0227 (0.0377)	0.0250 (0.0300)	0.0000 (0.0393)	0.0000 (0.0376)
rcmod	0.0388 (0.0359)	0.0000 (0.0336)	0.0267 (0.0395)	0.0000 (0.0386)
vmod	0.0194 (0.0247)	0.0278 (0.0306)	0.0000 (0.0330)	0.0000 (0.0235)
conj	0.0247 (0.0371)	0.0313 (0.0382)	0.0308 (0.0406)	0.0435 (0.0514)

assumptions suggest that some coefficient estimates should be interpreted with appropriate caution. The results above provided insights into the patterns of complex nominal structures across varieties, topics, and language proficiency levels, which laid the empirical foundation for addressing RQ3. The theoretical interpretation of these patterns will be presented in the Discussion section.

4 | Results

4.1 | Variation in Nominal Complexity Across English Varieties

The descriptive statistics for these nominal structures across the four sub-corpora are displayed in Table 4, including the median and standard deviation (SD) of the normalized frequencies, calculated based on as each type of dependent per nominal head. Kruskal—Wallis H tests were performed to examine potential differences in the distribution of complex nominal structures across the four English varieties. The results revealed significant differences in the relative frequencies of all eight complex nominal structures ($p < 0.05$ for all), indicating that patterns of nominal modification exhibit considerable variation depending on the context of English usage. Detailed results of the pairwise comparisons between the varieties are presented in Table 5. Additionally, boxplots visualizing the spread and central tendency of eight nominal structures across the four varieties (Figures 1 and 2), in conjunction with the descriptive statistics, highlight several noteworthy trends in nominal modification. Attributive adjectives are the most frequently occurring modifiers across all varieties, followed by prepositional phrases functioning as post-modifiers. Pre-modifiers such as nouns and possessives occur less frequently, each accounting for approximately half the occurrences of the leading two categories. Following these are verbal structure modifiers, relative clauses, adverbial modifiers, and noun phrase coordination.

The subsequent sections delve into a detailed analysis of the distribution of these eight nominal structures across the four varieties.

4.1.1 | Phrasal Modification and Nominal Phrase Coordination

In terms of nouns modified by adjectives, ESL learners employ this structure most frequently, followed by EFL learners and interpreter trainees. Interestingly, L1 speakers demonstrate significantly lower use of this structure compared to all three other groups.[1]

Moving to the use of prepositional phrases, interpreter trainees show a preference for using this structure as post-modifiers compared to all other groups. Conversely, L1 speakers use this structure the least, although the difference between L1 speakers and ESL/EFL learners is not statistically significant.[2]

Regarding the use of possessives as modifiers, EFL learners and interpreter trainees pattern together in using possessives significantly more than L1 speakers.[3]

Nouns used as modifiers appear more commonly in INT and ESL than in EFL and L1. The difference between EFL and the higher usage groups is statistically significant, suggesting EFL learners use statistically fewer nouns as modifiers compared to interpreter trainees and ESL learners.[4]

The analysis also reveals contrasting patterns in the use of adverbial modification across varieties. L1 English speakers demonstrate the highest frequency, differing significantly from

TABLE 5 | Kruskal—Wallis tests and post-hoc tests on the eight measures.

	Kruskal—Wallis H	*p*-value	L1 vs. ESL	L1 vs. EFL	L1 vs. INT	INT vs. ESL	INT vs. EFL	ESL vs. EFL
amod	34.713	0.000*	<*	<*	<*	<*	<	>
prep	103.720	0.000*	<	<	<*	>*	>*	>
poss	24.803	0.000*	<	<*	<*	>	<	<
nn	12.550	0.006*	<	>	>	<*	>	>*
advmod	16.819	0.001*	>	>*	>*	<	<	>
rcmod	67.579	0.000*	>*	>*	>*	>	<	<
vmod	96.203	0.000*	<	>	>*	<*	<*	>*
conj	28.038	0.000*	<	>	<*	>	>*	>

Note: *refers to statistically significant differences at $p < 0.05$; *p*-values reported are Bonferroni-adjusted.

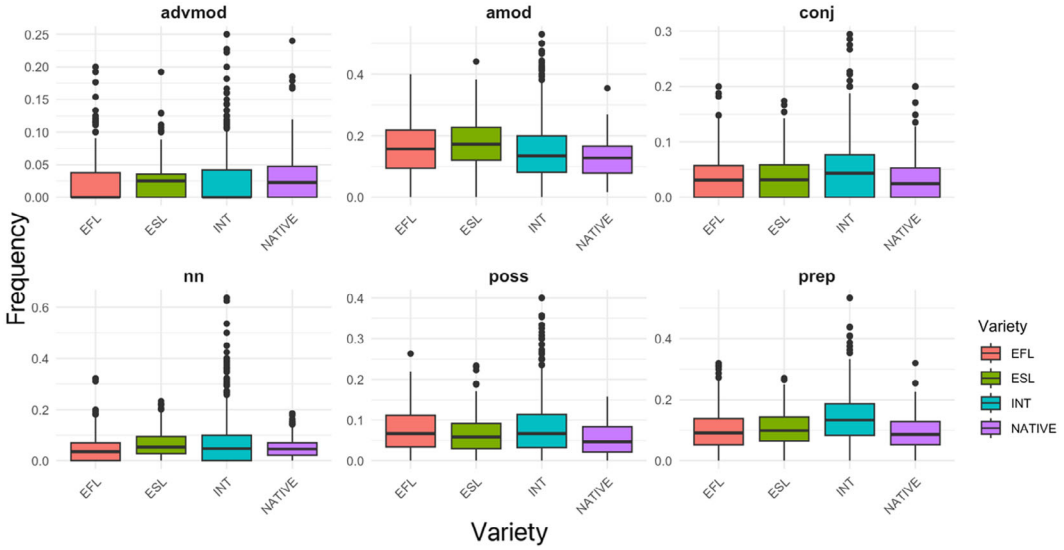

FIGURE 1 | Boxplots of indices of phrasal noun modification and nominal phrase coordination across varieties. [Color figure can be viewed at wileyonlinelibrary.com]

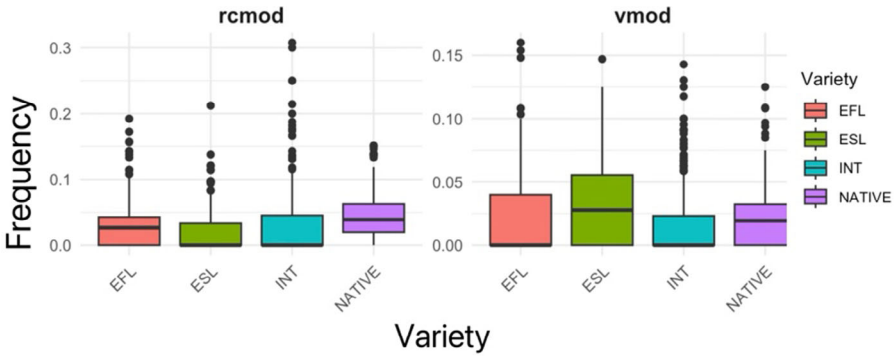

FIGURE 2 | Boxplots of indices of clausal noun modification across varieties. [Color figure can be viewed at wileyonlinelibrary.com]

the patterns observed for interpreter trainees who utilize them the least.[5]

Moving beyond single nominal phrases, the analysis reveals a significantly stronger preference for coordinating nominal phrases among interpreter trainees compared to all other groups, including L1 English speakers.[6]

4.1.2 | Clausal Modification

The analysis reveals a consistent pattern in relative clause usage, with all groups employing them slightly more frequently than non-finite verbal structure modifiers. However, L1 speakers demonstrate a stronger preference for this structure compared to the constrained language user groups. No statistically significant

TABLE 6 | Summary of models for eight nominal complexity measures.

		Fixed effect					Effect size	
	Parameters	**Estimate**	***SE***	***t***	***p***	**95% CI**	**Marginal R^2**	**Conditional R^2**
amod	(Intercept)	0.196	0.006	32.856	<0.001	[0.185, 0.208]	0.199	0.384
	Topic_Smoking	−0.060	0.005	−11.681	<0.001	[−0.071, −0.050]		
	CEFR_B2	0.000	0.009	−0.048	0.961	[−0.019, 0.018]		
	CEFR_C2	−0.039	0.008	−4.571	<0.001	[−0.055, −0.022]		
prep	(Intercept)	0.095	0.005	18.725	<0.001	[0.085, 0.104]	0.032	0.227
	Topic_Smoking	0.019	0.004	4.279	<0.001	[0.010, 0.028]		
	CEFR_B2	−0.008	0.008	−1.065	0.289	[−0.024, 0.007]		
	CEFR_C2	−0.012	0.007	−1.744	0.083	[−0.026, 0.001]		
poss	(Intercept)	0.085	0.004	21.666	<0.001	[0.078, 0.093]	0.088	0.077
	Topic_Smoking	−0.024	0.004	−6.841	<0.001	[−0.031, −0.017]		
	CEFR_B2	−0.004	0.006	−0.708	0.480	[−0.016, 0.008]		
	CEFR_C2	−0.019	0.005	−3.471	<0.001	[−0.030, −0.008]		
nn	(Intercept)	0.075	0.004	17.414	<0.001	[0.067, 0.084]	0.106	0.415
	Topic_Smoking	−0.033	0.003	−9.977	<0.001	[−0.040, −0.027]		
	CEFR_B2	−0.001	0.007	−0.201	0.841	[−0.015, 0.012]		
	CEFR_C2	−0.008	0.006	−1.189	0.236	[−0.020, 0.005]		
advmod	(Intercept)	0.025	0.003	9.083	<0.001	[0.020, 0.031]	0.015	0.075
	Topic_Smoking	0.005	0.003	1.904	0.058	[−0.000, 0.011]		
	CEFR_B2	−0.004	0.004	−1.037	0.302	[−0.012, 0.004]		
	CEFR_C2	0.005	0.004	1.325	0.187	[−0.002, 0.012]		
conj	(Intercept)	0.034	0.003	10.956	<0.001	[0.028, 0.041]	0.023	0.177
	Topic_Smoking	0.011	0.003	3.681	<0.001	[0.005, 0.016]		
	CEFR_B2	−0.006	0.005	−1.169	0.244	[−0.015, 0.004]		
	CEFR_C2	−0.005	0.004	−1.066	0.288	[−0.013, 0.004]		
rcmod	(Intercept)	0.020	0.003	7.139	<0.001	[0.015, 0.026]	0.072	0.118
	Topic_Smoking	0.012	0.003	4.088	<0.001	[0.006, 0.018]		
	CEFR_B2	0.003	0.004	0.833	0.406	[−0.005, 0.011]		
	CEFR_C2	0.018	0.004	4.961	<0.001	[0.011, 0.025]		
vmod	(Intercept)	0.033	0.002	14.025	<0.001	[0.028, 0.038]	0.077	0.225
	Topic_Smoking	−0.015	0.002	−6.845	<0.001	[−0.019, −0.011]		
	CEFR_B2	0.006	0.004	1.660	0.099	[−0.001, 0.013]		
	CEFR_C2	−0.004	0.003	−1.114	0.267	[−0.010, 0.003]		

Note: The CEFR baseline is B1, and the Topic baseline is part-time job.

difference in relative clause usage emerged among the three constrained groups.[7]

In contrast, the use of non-finite verbal structures as modifiers presents a different pattern. ESL learners and L1 speakers are more inclined toward this structure compared to EFL learners and interpreter trainees. Notably, interpreter trainees exhibit the least reliance on this structure, with their usage significantly lower than the other three groups.[8]

4.2 | Effects of Language Proficiency and Task Topic on Nominal Complexity

This section reports the influence of language proficiency and task topic on the complexity of nominal structures. By incorporating CEFR levels (B1, B2, C1, C2) and Topic (Part-time job—PTJ, Smoking—SMK) as predictors and including Student ID as a random effect, the analysis centers around three categories of nominal complexity: phrasal modification, clausal modification,

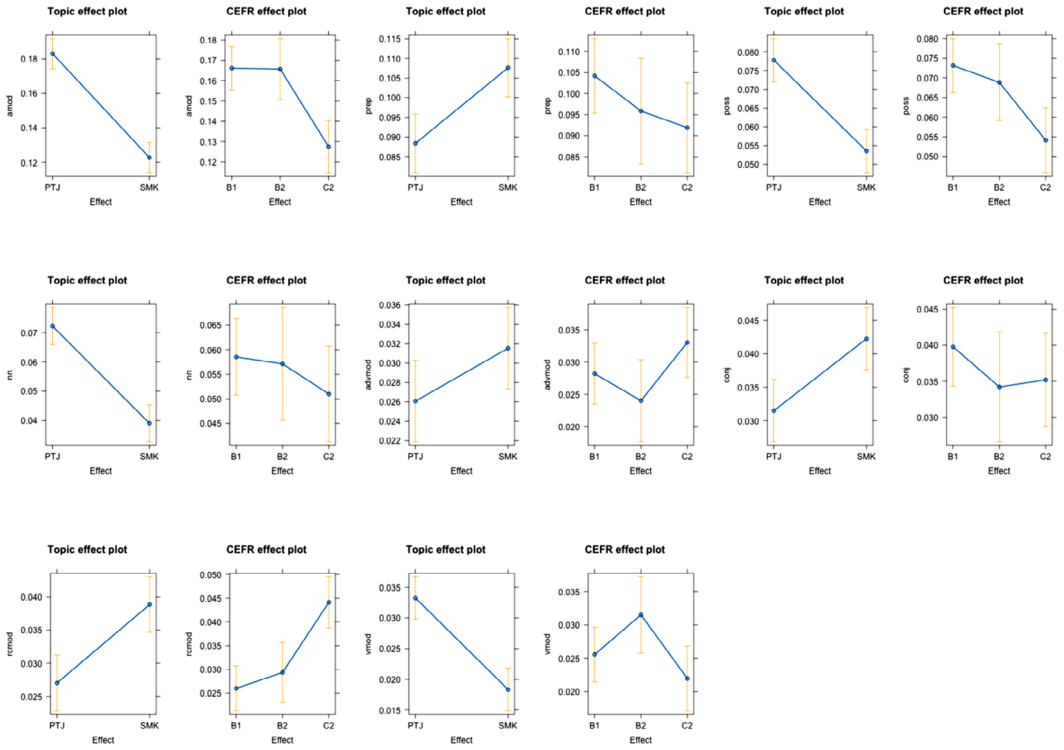

FIGURE 3 | Effects of English proficiency levels and topics in models for eight nominal complexity measures. [Color figure can be viewed at wileyonlinelibrary.com]

and noun phrase coordination. A summary of the modeling results is displayed in Table 6. The effect sizes of proficiency levels and task topics in these models are depicted in Figure 3

4.2.1 | Phrasal Modification and Nominal Phrase Coordination

The analysis reveals a noticeable effect of task topic on most phrasal structures. The SMK topic leads to a decreased use of nouns (Estimate = −0.033, $p < 0.001$) and possessives (Estimate = −0.024, $p < 0.001$) as modifiers but an increased use of prepositional phrases (Estimate = 0.019, $p < 0.001$) compared to the PTJ topic. Moreover, a significant effect of task topic is observed on noun phrase coordination, with the SMK topic associated with a higher frequency of coordinated nominal phrases compared to the other topics (Estimate = 0.011, $p < 0.001$). However, no significant topic effect is found on the use of adverbial modifiers, although the results are marginally significant (Estimate = 0.005, $p = 0.058$).

In contrast, language proficiency has a limited impact on phrasal modification. The only noteworthy effect is observed with possessive modifiers, where learners at the C2 level (L1) employ this structure less frequently than learners at lower proficiency levels (Estimate = −0.019, $p < 0.001$). However, no significant differences are found between B1 and B2 levels (both for ESL and EFL).

The marginal R-squared values (R^2m) indicate that fixed effects explained 19.9% of the variance in adjective modifiers and 10.6% in noun modifiers, while explaining considerably less variance in other phrasal modification structures (ranging from 1.5% to 8.8%). When including both fixed and random effects, the conditional R-squared values (R^2c) showed higher explanatory power across all structures: noun modifiers ($R^2c = 0.415$), adjective modifiers ($R^2c = 0.384$), prepositional modifiers ($R^2c = 0.227$), conjoining noun phrases ($R^2c = 0.177$), possessive modifiers ($R^2c = 0.077$), and adverbial modifiers ($R^2c = 0.075$). This pattern suggests that while fixed effects (English proficiency and topic) had moderate predictive power for some structures, participant-level variation captured by random effects contributed substantially to the overall model fit.

4.2.2 | Clausal Modification

Similarly, clausal modification is also significantly influenced by the task topic. The SMK topic results in more frequent use of relative clauses (Estimate = 0.012, $p < 0.001$) and fewer non-finite verbal structures (Estimate = −0.015, $p < 0.001$) as modifiers compared to the PTJ topic. With regard to language proficiency, C2 level learners (L1) exhibit a significantly higher preference for the use of relative clauses compared to learners at lower proficiency levels, with (Estimate = 0.018, $p < 0.001$) no differences observed between B1 and B2 levels.

The marginal R-squared values (R^2m) indicate that fixed effects explained approximately 7% of the variance in both relative clauses and non-finite verbal structures. When including both fixed and random effects, the conditional R-squared values (R^2c) revealed higher explanatory power: 11.8% for relative clauses and

22.5% for non-finite verbal structures. The notable increase in R^2c values, particularly for non-finite verbal structures, suggests that participant-level variation accounted for a substantial portion of the observed variance.

In summary, the analysis underscores the critical role of task topic in shaping the nominal choices of speakers across different language proficiency levels and varieties, affecting both phrasal and clausal modification as well as noun phrase coordination. Conversely, language proficiency level, as defined by the CEFR, does not generally affect the use of complex nominal structures, with the exception of possessive modifiers and relative clauses.

5 | Discussion

The present study has investigated the distribution of complex nominal structures across constrained and non-constrained varieties of English, revealing the modulation of speakers' English proficiency levels and the topics they discuss on these distributions. This section discusses the major findings from the analysis, paying particular attention to the shared and divergent patterns of nominal complexity in constrained varieties and examining whether such patterns can be attributed to the common and distinct constraints that characterize these varieties. By examining the interplay between the observed patterns and the underlying linguistic constraints, this section aims to provide a comprehensive understanding of the factors shaping the distribution of complex nominal structures in constrained Englishes.

5.1 | Shared Patterns of Nominal Complexity Indicate Common Constraints of Constrained Varieties to Varying Degrees

The analysis partially supports the hypothesis that constrained language varieties exhibit similar patterns in the distribution of complex nominal structures. Specifically, these varieties exhibit a higher frequency of adjective modifiers in noun modification compared to L1 English. Instead, they demonstrate a lower frequency of relative clauses and adverbial modifiers. This preference for phrasal modification (e.g., adjectives) over clausal modification (e.g., relative clauses) of constrained language users suggests that these varieties share common linguistic constraints.

The tendency for constrained speech to exhibit features typical of written English, like phrasal elaboration, as noted by Kruger and Van Rooy (2016), may be attributed to the prevalent use of English within educational and formal settings by L2 speakers and interpreters. The L2 English speakers and interpreter trainees in the current study, despite being intermediate to advanced learners of English, are less exposed to diverse registers and modes of English. This limited exposure may lead to more formal and less varied language production, as they are less familiar with the nuances of colloquial and informal English. Similar observations were also made in Gilquin's (2024) investigation of Hong Kong English and mainland China English produced by learners in a spoken task. The study found that L1 English speakers show a lower level of lexical sophistication compared to both constrained spoken varieties, which can be partly attributed to the more frequent use of colloquial vocabulary items by L1 speakers.

Several studies corroborate this finding across different language pairs and contexts (Chen et al. 2024; Ivaska and Bernardini 2020; Ivaska et al. 2022). For instance, Chen et al. (2024) observed a similar trend in written constrained language production, reporting a lower level of clausal structures and a higher level of complex nominals in English produced by Chinese writers and in Chinese-English translations. This pattern extends beyond the Chinese-English language pair, as evidenced by Ivaska and Bernardini (2020), who found noun phrase complexity to be a distinguishing factor between constrained and non-constrained Finnish across first/source languages and registers. Similarly, Ivaska et al. (2022) identified the choice between clausal and phrasal elaboration of information as a key differentiator between constrained and non-constrained languages, although the degree of difference varies across registers. These findings collectively suggest that the preference for phrasal modification over clausal modification is a shared characteristic of constrained English varieties, which can be attributed to the cognitive and contextual demands of specific communication situations faced by L2 speakers and interpreters. However, this tendency towards phrasal modification is not universally observed across all constrained language contexts. Liu et al. (2023) found that both interpreted and L2 English speech exhibited less complexity in subordination and complex nominals compared to L1 English speech. This divergence may be attributed to their study's focus on political discourse produced by professionals, a context potentially requiring more elaborate clausal structures to effectively convey complex ideas and arguments.

Additionally, the overuse of commonly employed structures, such as adjectives for noun modification, and the avoidance of less typical structures, such as verbal modifiers in interpreted speeches, reflects a normalization trend observed in both translation and L2 acquisition studies. This trend suggests that constrained language users may opt for more common and conventional expressions to ensure clarity and correctness, especially due to the higher communicative risks involved in constrained communication scenarios (Kruger and Van Rooy 2016).

However, some differences exist among the constrained varieties. ESL learners display a similar trend to L1 speakers in employing certain syntactic structures (e.g., more verbal structure modifiers and fewer possessive modifiers) compared to INT and EFL. This alignment suggests a convergence towards L1 speakers' norms in ESL, likely influenced by higher exposure and usage of English in diverse contexts by speakers of New Englishes, such as HK English. In contrast, EFL often exhibits more formal and monostylistic tendencies due to the restricted use of English in academic and formal settings (Gilquin 2024).

In conclusion, the shared patterns of nominal complexity in constrained language varieties indicate common constraints faced by L2 speakers and interpreters. The language usage settings and the normalization trend contribute to the observed similarities. However, differences among the constrained varieties, particularly between ESL and EFL, highlight the role of exposure and usage in shaping language production.

5.2 | Limited Effect of Language Proficiency While Prominent Effect of Task Topic Across Varieties

The analysis revealed intriguing findings regarding the impact of language proficiency and task topic on the use of complex nominal structures in language production by L1 and L2 speakers.

Regarding the effect of language proficiency, the study found no significant difference in the complex nominal structures used by speakers at CEFR levels B1 and B2. This finding aligns with Bao (2024) and Lan et al. (2019), both of whom observed minimal or no significant effect of proficiency on the employment of complex nominal phrases in both written and spoken English. However, the results contradict what has been proposed by Biber et al. (2011), which claims that the competence for formal writing (e.g., academic writing), characterized by features such as complex nominal phrases, is developed later in the learning process. The findings also differ from those of Bulté and Roothooft (2020), who found a significant effect of proficiency on L2 speech complexity, although their study operationalized noun phrasal elaboration differently and measured proficiency level using IELTS scores. The lack of differentiation among speakers at different proficiency levels might suggest that all participants, despite being at CEFR B1 to B2 levels, are still in relatively early stages of mastering complex nominal structures for spoken language production; thus, they may fail to adapt to the spoken mode of the production in this regard. As suggested by Bao (2024), the participants' proficiency levels might not have been sufficiently advanced to demonstrate notable differences in the complexity of language used across different modes, i.e., writing versus speaking.

In contrast to the limited effect of language proficiency, the task topic emerged as a significant factor influencing the use of complex nominal structures. This is consistent with previous studies by Lan et al. (2019), Yang et al. (2015), and Yoon (2017), which highlighted the substantial role of task topic in shaping the complexity of noun phrases and overall linguistic structures. The current study found that the repeated use of similar sequences within a given topic may contribute to this effect. A detailed examination of the data revealed that, when talking about smoking, learners frequently utilized noun phrases such as "environment for the customer" (prepositional phrases as modifiers) and "habit which can influence one's health" (relative clauses). Similarly, when discussing part-time jobs, noun phrases such as "part-time jobs" (adjective modifiers) and "their parents" (possessive modifiers) were commonly employed. These reoccurring sequences tend to lead to significantly higher use of certain nominal structures but the absence of others.

6 | Conclusion

The study investigated the distribution of complex nominal structures across constrained and non-constrained varieties of English, considering the influence of speakers' language proficiency and task topic. The findings suggest that constrained English speakers, including ESL learners, EFL learners, and interpreter trainees, demonstrate similar patterns in their use of complex nominal structures, with a preference for phrasal over clausal modification. This tendency towards a "bookish" and formal style likely serves as a compensatory strategy for limited exposure to informal L1 English. Notably, while language proficiency did not significantly influence nominal complexity, the task topic emerged as a crucial factor in shaping the nominal complexity of English varieties.

The findings of this study offer valuable insights into both theoretical frameworks and pedagogical practices within translation studies, second language acquisition, and bilingualism research. From a theoretical standpoint, the study bridges these disciplines by adopting a unified framework that treats interpreted and L2 English speech as constrained languages. This approach underscores the importance of examining granular indices of grammatical complexity rather than treating them as a monolithic construct. Such detailed analysis can refine theoretical models of language production in bilingual and multilingual contexts, accounting for the interplay between language proficiency, task demands, and communicative context. Pedagogically, the findings emphasize the necessity for curriculum designs that provide L2 English speakers and interpreter trainees with extensive exposure to diverse complex nominal structures. Incorporating authentic texts that showcase varied nominal complexities, along with explicit instruction and targeted practice, can facilitate a deeper understanding of complex nominal structures across different communicative contexts. These strategies are essential for developing a comprehensive grasp of how these structures function in real-world scenarios.

Further research should address several limitations to enhance the reliability and generalizability of the findings. First, the inconsistent topic categorization across sub-corpora may have introduced confounding variables. To control topic-related variability, further studies should adopt an experimental design in which all participants speak or write on identical topics or prompts, providing more robust and comparable results. Second, the potential influence of speakers' first language (L1) on their use of complex nominal structures was not thoroughly examined. Future research should consider both the source language for interpreting corpora and comparable L1 production for EFL/ESL corpora to investigate transfer effects from the speakers' L1.

A third limitation concerns the substantial unexplained variance captured by random effects in our analysis. Our results showed that participant-level variation accounts for a considerable proportion of variance in nominal complexity measures, an observation that aligns with previous research. For instance, Kajzer-Wietrzny (2022) found that individual speaker effects were the primary source of variation in the frequency of cohesive devices. This finding resonates with translanguaging theory (Wei 2018), which suggests that the conventional separation of languages may not accurately reflect how language users dynamically draw upon their linguistic resources. The significant role of individual variation highlights the need to examine speaker-specific factors more systematically in future research, including individual linguistic repertoires, language learning histories, and cognitive processing styles.

Ethics Statement

This article does not contain any studies with human participants performed by any of the authors.

Conflicts of Interest

The authors declare no conflict of interests.

Data Availability Statement

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

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1111/ijal.12725.

Permission to reproduce material from other sources

Part of the data is obtained online from publicly available sources and is cited appropriately.

Endnotes

[1] L1 versus INT: z = 2.765837, *p.adj* = 0.0170*; L1 versus ESL: z = 5.548923, *p.adj* = 0.0000*; L1 versus EFL: z = 3.716074, *p.adj* = 0.0006*; INT versus EFL: z = 1.932112, *p.adj* = 0.1600; INT versus ESL: z = 4.238293, *p.adj* = 0.0001*; EFL versus ESL: z = -1.823662, *p.adj* = 0.2046.

[2] L1 versus INT: z = 8.061691, *p.adj* = 0.0000*; L1 versus ESL: z = 1.841016, *p.adj* = .1969; L1 versus EFL: z = 0.957034, *p.adj* = 1.0000; INT versus EFL: z = -6.792784, *p.adj* = 0.0000*; INT versus ESL: z = -5.680513, *p.adj* = 0.0000*; EFL versus ESL: z = -0.879551, *p.adj* = 1.0000.

[3] L1 versus INT: z = 4.548500, *p.adj* = 0.0000*; L1 versus ESL: z = 1.931029, *p.adj* = 0.1604; L1 versus EFL: z = 3.963643, *p.adj* = 0.0002*; INT versus EFL: z = 0.475265, *p.adj* = 1.0000; INT versus ESL: z = -2.082268, *p.adj* = 0.1120; EFL versus ESL: z = 2.022425, *p.adj* = 0.1294.

[4] L1 versus INT: z = 0.894985, *p.adj* = 1.0000; L1 versus ESL: z = 1.891768, *p.adj* = .1756; L1 versus EFL: z = -1.497467, *p.adj* = .4028; INT versus EFL: z = -2.771986, *p.adj* = .0167*; INT versus ESL: z = 1.492514, *p.adj* = .4067; EFL versus ESL: z = -3.372246, *p.adj* = .0022*.

[5] L1 vs. INT: z = -3.991305, *p.adj* = .0002*; L1 versus ESL: z = -1.792994, *p.adj* = .2189; L1 versus EFL: z = -2.559165, *p.adj* = .0315*; INT versus EFL: z = 0.739195, *p.adj* = 1.0000; INT versus ESL: z = 1.703229, *p.adj* = .2656; EFL versus ESL: z = -0.762330, *p.adj* = 1.0000.

[6] L1 versus INT: z = 3.984756, *p.adj* = .0002*; L1 versus ESL: z = 1.277037, *p.adj* = .6048; L1 versus EFL: z = -0.090550, *p.adj* = 1.0000; INT versus EFL: z = -4.066701, *p.adj* = 0.0001*; INT versus ESL: z = -2.345935, *p.adj* = 0.0569; EFL versus ESL: z = -1.360733, *p.adj* = 0.5208.

[7] L1 versus INT: z = -7.632749, *p.adj* = 0.0000*; L1 versus ESL: z = -7.114105, *p.adj* = 0.0000*; L1 versus EFL: z = -4.681366, *p.adj* = 0.0000*; INT versus EFL: z = 1.681148, *p.adj* = 0.2782; INT versus ESL: z = -1.379841, *p.adj* = 0.5029; EFL versus ESL: z = 2.420544, *p.adj* = 0.0465.

[8] L1 versus INT: z = -5.611025, *p.adj* = 0.0000*; L1 versus ESL: z = 2.339514, *p.adj* = .0579; L1 versus EFL: z = -0.376753, *p.adj* = 1.0000; INT versus EFL: z = 5.091929, *p.adj* = 0.0000*; INT versus ESL: z = 8.509671, *p.adj* = 0.0000*; EFL versus ESL: z = -2.702652, *p.adj* = 0.0206*.

References

Ansarifar, A., H. Shahriari, and R. Pishghadam. 2018. "Phrasal Complexity in Academic Writing: A Comparison of Abstracts Written by Graduate Students and Expert Writers in Applied Linguistics." *Journal of English for Academ* 31: 58–71. https://doi.org/10.1016/j.jeap.2017.12.008.

Bao, G. 2024. "Differential Effects of EFL Proficiency and Mode on Noun Phrase Use." *System* 121: 103244. https://doi.org/10.1016/j.system.2024.103244.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2014. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67, no. 1: 1–48. https://doi.org/10.48550/arXiv.1406.5823.

Biber, D. 1988. *Variation Across Speech and Writing* 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511621024.

Biber, D., and B. Gray. 2010. "Challenging Stereotypes About Academic Writing: Complexity, Elaboration, Explicitness." *Journal of English for Academ* 9, no. 1: 2–20. https://doi.org/10.1016/j.jeap.2010.01.001.

Biber, D., B. Gray, and K. Poonpon. 2011. "Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development?" *TESOL Q* 45, no. 1: 5–35. https://doi.org/10.5054/tq.2011.244483.

Biber, D., B. Gray, S. Staples, and J. Egbert. 2020. "Investigating Grammatical Complexity in L2 English Writing Research: Linguistic Description Versus Predictive Measurement." *Journal of English for Academic Purposes* 46: 100869. https://doi.org/10.1016/j.jeap.2020.100869.

Biber, D., T. Larsson, and G. R. Hancock. 2023. "The Linguistic Organization of Grammatical Text Complexity: Comparing the Empirical Adequacy of Theory-based Models." *Corpus Linguistics and Linguistic Theory* 20, no. 2: 347–373. https://doi.org/10.1515/cllt-2023-0016.

Bulté, B., and H. Roothooft. 2020. "Investigating the Interrelationship Between Rated L2 Proficiency and Linguistic Complexity in L2 Speech." *System* 91, 102246. https://doi.org/10.1016/j.system.2020.102246.

Chesterman, A. 2004. "Hypotheses About Translation Universals." In *Benjamins Translation Library*, edited by G. Hansen, K. Malmkjaer, and D. Gile, 50, 1–13. John Benjamins Publishing Company. https://doi.org/10.1075/btl.50.02che.

Cho, M. 2018. "Task Complexity, Modality, and Working Memory in L2 Task Performance." *System* 72: 85–98. https://doi.org/10.1016/j.system.2017.10.010.

De Clercq, B. 2015. "The Development of Lexical Complexity in Second Language Acquisition. A Cross-Linguistic Study of L2 French and English." *Eurosla yearbook*, edited by L. Roberts, K. McManus, N. Vanek, and D. Trenkic (Vol. 15), 69–94. John Benjamins. https://doi.org/10.1075/eurosla.15.03dec.

De Clercq, B., and A. Housen. 2017. "A Cross-linguistic Perspective on Syntactic Complexity in L2 Development: Syntactic Elaboration and Diversity." *The Modern Language Journal* 101, no. 2: 315–334. https://doi.org/10.1111/modl.12396.

De Sutter, G., and M. A. Lefer. 2020. "On the Need for a New Research Agenda for Corpus-Based Translation Studies: A Multi-Methodological, Multifactorial and Interdisciplinary Approach." *Perspectives (Gerontological Nursing Association (Canada))* 28, no. 1: 1–23. https://doi.org/10.1080/0907676X.2019.1611891.

Gan, Z. 2012. "Complexity Measures, Task Type, and Analytic Evaluations of Speaking Proficiency in a School-Based Assessment Context." *Language Assessment Quarterly* 9, no. 2: 133–151. https://doi.org/10.1080/15434303.2010.516041.

Gilquin, G. 2024. "Lexical Use in Spoken New Englishes and Learner Englishes: The Effects of Shared and Distinct Communicative constraints." In *Constraints on Language Variation and Change in Complex Multilingual Contact Settings*, edited by B. van Rooy and H. Kotze, 120–152. John Benjamins.

Granger, S. 2015. "Contrastive Interlanguage Analysis: A Reappraisal." *International Journal of Learner Corpus Research* 1, no. 1: 7–24. https://doi.org/10.1075/ijlcr.1.1.01gra.

Halliday, M. A. K. 1990. *Spoken and Written Language*. Oxford University Press.

Hayashi, F. 2000. *Econometrics*. Princeton University Press.

Iwashita, N., A. Brown, T. McNamara, and S. O'Hagan. 2008. "Assessed Levels of Second Language Speaking Proficiency: How Distinct?" *Applied Linguistics* 29, no. 1: 24–49. https://doi.org/10.1093/applin/amm017.

Ishikawa, S. 2023. *The Incnale Guide: An Introduction to a Learner Corpus Study on Asian Learners' L2 English*. Routledge.

Ivaska, I., and S. Bernardini. 2020. "Constrained Language Use in Finnish: A Corpus-Driven Approach." *Nordic Journal of Linguistics* 43, no. 1: 33–57. https://doi.org/10.1017/S0332586520000013.

Ivaska, I., A. Ferraresi, and S. Bernardini. 2022. "Syntactic Properties of Constrained English: A Corpus-Driven Approach." In *Extending the Scope of Corpus-Based Translation Studies*, edited by S. Granger and M. A. Lefer, 133–157. Bloomsbury Academic.

Jucker, A. H. 1992. *Social Stylistics: Syntactic Variation in British Newspapers*. Mouton de Gruyter.

Kajzer-Wietrzny, M. 2018. "Interpretese vs. Non-native Language Use: The Case of Optional *That*." In *Making Way in Corpus-based Interpreting Studies*, edited by M. Russo, C. Bendazzoli, and B. Defrancq, 97–113. Springer.

Kajzer-Wietrzny, M. 2022. "An Intermodal Approach to Cohesion in Constrained and Unconstrained Language." *Target International Journal of Translation Studies* 34, no. 1: 130–162. https://doi.org/10.1075/target.19186.kaj.

Kajzer-Wietrzny, M., and I. Ivaska. 2020. "A Multivariate Approach to Lexical Diversity in Constrained Language." *Across Languages and Cultures* 21, no. 2: 169–194. https://doi.org/10.1556/084.2020.00011.

Kormos, J. 2014. "Differences Across Modalities of Performance: An Investigation of Linguistic and Discourse Complexity in Narrative tasks." In *Task-Based Language Learning–Insights from and for L2 Writing*, edited by H. Byrnes and R. M. Manchón, 193–217. John Benjamins. https://doi.org/10.1075/tblt.7.08kor.

Kotze, H. 2022. "Translation as Constrained Communication: Principles, Concepts and Methods." In *Extending the Scope of Corpus-Based Translation Studies*, edited by S. Granger and M. A. Lefer, 67–97. Bloomsbury Publishing. http://www.bloomsburycollections.com/book/extending-the-scope-of-corpus-based-translation-studies/ch3-translation-as-constrained-communication/.

Kruger, H., and B. Van Rooy. 2016. "Constrained Language: A Multidimensional Analysis of Translated English and a Non-Native Indigenised Variety of English." *English World-Wide* 37, no. 1: 26–57. https://doi.org/10.1075/eww.37.1.02kru.

Kruger, H., and B. van Rooy. 2018. "Register Variation in Written Contact Varieties of English: A Multidimensional Analysis." *English World-Wide* 39, no. 2: 214–242. https://doi.org/10.1075/eww.00011.kru.

Kyle, K., and S. A. Crossley. 2018. "Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices." *Modern Language Journal* 102, no. 2: 333–349. https://doi.org/10.1111/modl.12468.

Lan, G., K. Lucas, and Y. Sun. 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.

Lanstyák, I., and P. Heltai. 2012. "Universals in Language Contact and Translation." *Across Languages and Cultures* 13, no. 1: 99–121. https://doi.org/10.1556/Acr.13.2012.1.6.

Liu, Y., A. K. F. Cheung, and K. Liu. 2023. "Syntactic Complexity of Interpreted, L2 and L1 Speech: A Constrained Language Perspective." *Lingua* 286: 103509. https://doi.org/10.1016/j.lingua.2023.103509.

Lu, X. 2010. "Automatic Analysis of Syntactic Complexity in Second Language Writing." *International Journal of Corpus Linguistics* 15, no. 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, no. 2: 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x.

Lu, X. 2017. "Automated Measurement of Syntactic Complexity in Corpus-Based L2 Writing Research and Implications for Writing Assessment." *Language Testing* 34, no. 4: 493–511. https://doi.org/10.1177/0265532217710675.

Norris, J. M., and L. Ortega. 2009. "Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity." *Applied Linguistics* 30, no. 4: 555–578. https://doi.org/10.1093/applin/amp044.

Quirk, R., R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman.

Schilk, M., and S. Schaub. 2016. "Noun Phrase Complexity Across Varieties of English: Focus on Syntactic Function and Text Type." *English World-Wide* 37, no. 1: 58–85. https://doi.org/10.1075/eww.37.1.03sch.

Seedhouse, P., A. Harris, R. Naeb, and E. Üstünel. 2014. "The Relationship Between Speaking Features and Band Descriptors: A Mixed Method Study." *IELTS Research Reports Online Series* 2: 1–30.

Shlesinger, M., and N. Ordan. 2012. "More Spoken or More Translated? Exploring a Known Unknown of Simultaneous Interpreting." *Target* 24, no. 1: 43–60.

Ulrych, M., and A. Murphy. 2008. Descriptive Translation Studies and the Use of Corpora: Investigating Mediation Universals. In *Corpora for University Language Teachers*, edited by C. Taylor Torsello, K. Ackerley, and E. Castello, 141–166. Peter Lang.

Wei, L. 2018. "Translanguaging as a Practical Theory of Language." *Applied Linguistics* 39, no. 1: 9–30. https://doi.org/10.1093/applin/amx039.

Yang, W., X. Lu, and S. C. Weigle. 2015. "Different Topics, Different Discourse: Relationships Among Writing Topic, Measures of Syntactic Complexity, and Judgments of Writing Quality." *Journal of Second Language Writing* 28: 53–67. https://doi.org/10.1016/j.jslw.2015.02.002.

Yoon, H. J. 2017. "Linguistic Complexity in L2 Writing Revisited: Issues of Topic, Proficiency, and Construct Multidimensionality." *System* 66: 130–141. https://doi.org/10.1016/j.system.2017.03.007.

Anonymous author reference

Chen, J., D. Li, and K. Liu. 2024. "Unraveling Cognitive Constraints in Constrained Languages: A Comparative Study of Syntactic Complexity in Translated, EFL, and Native Varieties." *Language Sciences* 102: 101612. https://doi.org/10.1016/j.langsci.2024.101612.

Yao, Y., D. Li, Y. Huang, and Z. Sang. 2024. "Linguistic Variation in Mediated Diplomatic Communication: A Full Multi-Dimensional Analysis of Interpreted Language in Chinese Regular Press Conferences." *Humanities and Social Sciences Communications* 11, no. 1: 1409. https://doi.org/10.1057/s41599-024-03967-6.