Available online at www.sciencedirect.com

ScienceDirect

Lingua 286 (2023) 103509

Lingua

www.elsevier.com/locate/lingua

Syntactic complexity of interpreted, L2 and L1 speech: A constrained language perspective

Yi Liu, Andrew K.F. Cheung, Kanglong Liu *

Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

Received 13 October 2022; revised 15 February 2023; accepted in revised form 15 February 2023;

Abstract

This study investigates the differences in syntactic complexity among three language varieties: interpreted speech, nonnative English (L2) speech, and native English (L1) speech. The examination was conducted through the evaluation of 14 measures across five subconstructs: the length of the production unit, the amount of subordination, the amount of coordination, phrasal sophistication, and overall sentence complexity. We used a self-compiled comparable corpus of these three language varieties and tested the simplification hypothesis under the framework of constrained language. Our results showed that the two spoken constrained varieties, interpreted speech and L2 speech, had significantly lower scores on most of the syntactic complexity measures compared to non-constrained L1 speech. However, there was no consistent pattern between the two constrained varieties. Specifically, interpreted speech had longer language units and more coordination than L2 speech, which contained more subordination. Overall, this study provides new insights for simplification research by examining syntactic complexity measures from a constrained language perspective.

Keywords: Syntactic complexity; Constrained languages; Simplification; Interpreting; L2SCA

1. INTRODUCTION

Over the past few years, researchers in fields such as translation and interpreting, second language acquisition (SLA), and World Englishes have been investigating translated English and non-native varieties of English, which are often referred to as constrained languages (Kotze, 2022; Lanstyák and Heltai, 2012). These scholars are interested in understanding the unique features that distinguish these two language varieties from native English (Ivaska et al., 2022; Kajzer-Wietrzny and Ivaska, 2020; Kotze, 2022; Kruger and Van Rooy, 2016a; Lanstyák and Heltai, 2012). Simplification, which is one of the potential features that has received significant attention, has been widely studied in corpus-based translation and interpreting research for several decades (Baker, 1993, 1996). According to Baker (1996:176), simplification is the process by which translators "subconsciously simplify the language or message or both." Previous research on simplification has mainly focused on examining the regularities and patterns of translated

* Corresponding author.
E-mail addresses: crystal-yi.liu@connect.polyu.hk (Y. Liu), andrew.cheung@polyu.edu.hk (A.K.F. Cheung), kl.liu@polyu.edu.hk (K. Liu).

https://doi.org/10.1016/j.lingua.2023.103509
0016-7037/

and interpreted texts compared to non-translated and non-interpreted texts through a number of linguistic features. In a similar vein, simplification measures have also been used in SLA to evaluate the language performance of language learners, using the output of native speakers as the baseline (Larsen-Freeman, 2009; Lu, 2011; Norris and Ortega, 2009).

Translation and interpreting studies (TIS) and SLA are two separate fields that study linguistic deviation. In SLA, this means examining the ways in which non-native second languages (L2s) differ from native languages (L1s). In TIS, this involves analyzing the differences between translated/interpreted language and non-translated native language (Ivaska et al., 2022). Both translated languages and L2s are believed to be influenced by similar cognitive and social constraints (Ivaska et al., 2022; Kruger and Van Rooy, 2016a). Despite these shared research goals and assumptions, translation and interpreting and SLA have historically been studied separately from each other. However, Lanstyák and Heltai (2012) introduced the concept of "constrained communication," which aims to investigate the similarities in linguistic tendencies between L2s and translation and interpreting due to processing constraints. Most existing research on constrained communication seeks to determine whether there are shared linguistic tendencies between these two language varieties as a result of these constraints.

There has been relatively little research on interpreted language from the perspective of constrained languages. Translation is often considered to be "the most extreme end of the bilingual activation continuum," as the process of translation requires rapid bidirectional switching between source and target texts, leading to cognitive constraints on the language processing of the translator (Kruger and Van Rooy, 2016b:121). Spoken-constrained varieties such as simultaneous interpreting (SI) are even more constrained in terms of cognitive load, as SI interpreters are required to produce corresponding output instantly while listening to the source speech with a time lag of only a few seconds (Christoffels and De Groot, 2004; Gerver, 1976; Kroll and De Groot, 2005; Liang et al., 2019; Seeber, 2011). Therefore, it is important to include interpreting in the constrained language research paradigm, as it may exhibit different linguistic patterns from written translation (Bernardini et al., 2016; Kajzer-Wietrzny, 2015; Sandrelli and Bendazzoli, 2005). Similarly, spoken L2 varieties, which are not subject to "intermediate intervention" (Kajzer-Wietrzny and Grabowski, 2021:150), should also be included in the analysis, as they may reveal unique linguistic patterns.

When analyzing interpreted speech, it is important to consider the directionality of interpreting. Most existing studies on interpreting have focused on interpreting in the first or "A" language (Kajzer-Wietrzny and Ivaska, 2020), as this is the preferred direction for most researchers and interpreters (Bartłomiejczyk, 2006). However, interpreting from the first or "A" language into the second or "B" language (L1–L2 interpreting) is also a common practice in private markets and international institutions (Gile, 2009). Nevertheless, empirical research on L1–L2 interpreting has received relatively little attention. In addition, interpreting into the second or "B" language (L2) is considered to be particularly constrained, as it involves the constraints of both interpretation (mediation) and foreign language use (Kajzer-Wietrzny and Grabowski, 2021).

As stated above, previous empirical studies on L2 and translation-mediated language have largely focused on written varieties, while little effort has been made to examine the simplification features in spoken L2 varieties and interpreting (Kajzer-Wietrzny and Ivaska, 2020). This study aims to address this gap by comparing three language varieties—interpreted speech, L2 speech, and L1 speech—within the unified framework of constrained communication. Specifically, this study focuses on L1–L2 interpreting, which is subject to twofold constraints. By systematically analyzing interpreted output and L2 spoken language, this study aims to gain insight into how these language productions are affected by shared and non-shared constraints.

Simplification features have been studied in both interpreting studies and SLA using various measurements. In interpreting studies, lexical indices such as lexical density, the type-token ratio, core vocabulary, and list head coverage (Laviosa, 1998b) have been commonly used to compare interpreted and native speech, interpreting and translation texts, and interpreting outputs in different modes (Dayter, 2018; Ferraresi et al., 2018; Kajzer-Wietrzny, 2015; Lv and Liang, 2019; Shlesinger and Ordan, 2012). Previous research has suggested that interpreted language is simpler than native speech and translated language when the target language is English (Bernardini et al., 2016; Ferraresi et al., 2018; Kajzer-Wietrzny, 2015). Research on SLA, on the other hand, examines simplification at both the lexical and syntactic levels using various syntactic complexity metrics. Vercellotti (2019) argued that syntactic complexity metrics can be a reliable way to assess the complexity of the language used by L2 speakers, as learners do not have the time to review or revise their language production or expand a message with modifying words when speaking orally. Therefore, syntactic complexity, especially in speech production, can closely reflect the cognitive load of L2 users. In addition, syntactic complexity is closely related to issues of cognitive resource management and the language processing mechanisms underlying SI (Hild, 2011; Seeber and Kerzel, 2012). A systematic investigation of syntactic complexity can provide empirical evidence for understanding the interpreting process, which is governed by the interplay of cognitive constraints and language representations (Jiang and Jiang, 2020).

This study aims to investigate the simplification hypothesis from the perspective of spoken constrained communication using syntactic complexity as an indicator. Our primary working hypothesis is that interpreters and L2 speakers will produce syntactically simpler structures than L1 speakers due to increased processing constraints. Furthermore, we predict that interpreters and L2 speakers may exhibit different patterns of syntactic complexity in their language production due to the different types of constraints that they encounter.

The structure of this paper is as follows: In Section 2, we discuss the measurement of syntactic complexity and provide an overview of relevant research on simplification in interpreting studies and constrained language. In Section 3, we describe the materials and methodology used in this study. The results of our analysis of syntactic complexity in the three speech varieties are presented in Section 4. In Section 5, we discuss the implications of these results. Finally, we conclude the paper in Section 6.

2. RELATED WORK

2.1. Measuring syntactic complexity

Syntactic complexity refers to the range and sophistication of forms that appear in language production (Ortega, 2003). SLA researchers have investigated how syntactic complexity can vary based on various external factors, such as linguistic proficiency (Bulté and Housen, 2012; Kyle and Crossley, 2018; Lu, 2011; Norris and Ortega, 2009), task modality (i.e., oral, written) (Biber et al., 2011; Kuiken and Vedder, 2011), the task type and genre (Biber et al., 2016; Lambert and Kormos, 2014), and the learner's L1 and L2 (Bernardini and Granfeldt, 2019; De Clercq and Housen, 2017; Lu and Ai, 2015). However, there is a lack of consensus on how to conceptualize and operationalize syntactic complexity, as a variety of measures have been used in this field of research (Pallotti, 2015).

Syntactic length is often used as an indicator of syntactic complexity, as it is believed that changes in length reflect underlying structural changes (De Clercq and Housen, 2017). In general, longer units are considered more complex than shorter units. Additionally, complex sentences are often derived from the combination of simple sentences through coordination or subordination. Thus, the frequency and depth of embedding are often considered a measure of complexity. Some linguistic units, such as coordinate phrases and complex nominals, are also thought to be more complex than others (Lu, 2010). Therefore, a high occurrence of these units can also be associated with increased syntactic complexity.

In recent years, there has been a call for the use of fine-grained measures of syntactic complexity to more accurately capture the complexity of syntactic structures (Biber et al., 2020; Kyle and Crossley, 2018). Fine-grained measures focus on specific features or aspects of syntax, such as dependencies and the use of certain constructions, while holistic measures take a broader view and consider a wider range of features. Both types of measures have their strengths and weaknesses and can be effectively applied depending on the research context. Holistic measures, such as the Biber Tagger (Biber et al., 1999), Coh-Metrix (McNamara et al., 2014), the L2 Syntactic Complexity Analyzer (L2SCA) (Lu, 2010), and the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASC) (Kyle, 2016), are particularly useful for examining the general trend of syntactic complexity in language production rather than specific language features. In fact, Lu (2017) found that the syntactic measures in the Biber Tagger, Coh-Metrix, and the L2SCA were all positively correlated with holistic writing measures, regardless of how the complexity measures were operationalized. In the current study, which aims to investigate the simplification hypothesis in interpreted and L2 productions, it is thought that holistic syntactic measures will be more effective in capturing the nature of syntactic complexity.

In this study, we chose to use the L2SCA to measure syntactic complexity due to its accessibility and reliability (Lu, 2010, 2017). The L2SCA assesses 14 syntactic complexity measures across five constructs: the length of the production unit, the amount of subordination, the amount of coordination, phrasal sophistication, and overall sentence complexity (Lu and Ai, 2015). While these measures may be partially redundant, we have chosen to include all 14 to provide a comprehensive understanding of syntactic complexity in the current study, despite the potential for superfluous information. The L2SCA has been widely used to examine writing by L2 learners (Casal and Lee, 2019; Khushik and Huhta, 2020; Lu, 2010, 2017; Lu and Ai, 2015) and has been shown to have high reliability in distinguishing language production between L1 and L2 speakers as well as between L2 speakers of different proficiency levels (Khushik and Huhta, 2020).

Notably, the reliability of the L2SCA may be decreased when applied to language production with various types of grammatically incomplete sentences (Lu, 2010). To mitigate this potential issue, we have removed elements such as false starts, repetitions, and filled pauses from our corpora. Despite this potential limitation, the L2SCA has been used successfully in studies on spoken corpora. For example, Hwang et al. (2020) applied all 14 syntactic complexity measures to written and spoken data produced by beginning-level English as a Foreign Language (EFL) child learners and

found that young EFL learners used less complex syntactic structures due to their immature linguistic and cognitive development. Similarly, Park (2022) analyzed monologues of Korean EFL undergraduates using the L2SCA and found that the spoken output of EFL learners was overrepresented with coordination, regardless of age. These studies demonstrate that the L2SCA can be effectively used to study interpreted language, as long as unwanted textual noise is properly cleaned.

2.2. Simplification in interpreting studies

The findings on simplification in interpreting research have been somewhat inconclusive. Most studies in this area have been conducted in the European context, comparing speeches delivered in the European Parliament and their corresponding interpretations. These studies have largely focused on lexical measures, such as lexical density, core vocabulary, and the list head, to examine simplification patterns. For example, Sandrelli and Bendazzoli (2005) used the European Parliament Interpreting Corpus (EPIC) to investigate whether the simplification patterns found in written translation (Laviosa, 1998a) could also be confirmed in interpreted language. Their study found that the simplification patterns identified by Laviosa were largely confirmed in interpreted English but varied depending on the language combination. Kajzer-Wietrzny (2012) used lexical density, core vocabulary, and the list head as simplification metrics to examine interpreting in different language pairs and source-speech modes and found that simplification existed only in the feature of list heads and not in other metrics. She also cautioned that the results might be affected by the language combination and mode of delivery of the source speech (i.e., spontaneous spoken speech or read-out speech). In a later study, Kajzer-Wietrzny (2015) confirmed that interpreted language is lexically simpler than native English speech in terms of lexical density and the proportion of high-frequency words. However, she found that interpreted language showed different repetition patterns that varied depending on the language combination and mode of delivery of the source speech.

When comparing translation and interpretation of the same source texts, Bernardini et al. (2016) found that the mediation process (translation and interpreting) reduces the lexical complexity of language production in both language directions (English and Italian), with interpreters simplifying the message more than translators. In addition, they found that translated English tended to be characterized by more lexical simplification, while translated Italian showed more lexico-syntactic simplification. Ferraresi et al. (2018) also compared the lexical features of translation and interpretation into English from different source languages using a multilingual intermodal corpus and found that the source language may have a greater impact on simplification levels than the mediation mode. Overall, these studies suggest that the simplification patterns in translation and interpretation may vary depending on several factors, including the language combination, mode of delivery, and source language. For studies focusing on language pairs other than European languages, Lv and Liang (2019) used corpus-based methods to analyze Chinese-English interpreting. They regarded lexical simplification as a dependent variable of the cognitive load in two different modes of interpreting (i.e., SI vs consecutive interpreting (CI)). CI was found to be more simplified than SI in various simplification metrics, including information density, lexical repetitiveness, and lexical sophistication. Based on these findings, the researchers suggested that the cognitive load of CI might be as high as, or even higher than, that of SI.

The advancement of methodology in this area of research has been slow, with many studies replicating the methodology proposed by Laviosa (1998a) and examining lexical simplification. These studies have generally confirmed that interpreted language is more simplified than native language at the lexical level. However, syntactic simplification metrics have received less attention in this line of inquiry, with relatively few studies focusing on this aspect of interpreted language. Overall, the research in this area has largely focused on lexical simplification, with syntactic simplification receiving less attention.

Compared to lexical features, syntactic complexity is a more robust measurement for investigating simplification because it can provide deeper insights into the interpreting process. This measurement is based on the complexity of various categories of syntactic structures (Xu and Li, 2021). Previous research has shown that the “morphosyntactic asymmetry” between source and target texts (Seeber and Kerzel, 2012:228) can increase the cognitive workload of interpreters and that syntactic complexity is a valid indicator of the cognitive load experienced by simultaneous interpreters (Hyönä et al., 1995; Seeber and Kerzel, 2012; Tommola and Niemi, 1986). Therefore, it is believed that syntactic complexity can be used in our study as an indicator to determine the degree to which interpreted language differs from other language varieties such as L2 speech and native speech.

2.3. Empirical investigation into constrained language

Under the unified framework of constrained language, researchers have begun to view translation/interpreting and L2 output as constrained languages that may display some common linguistic features (Kruger and Van Rooy,

2016a). Constrained language, defined as “the language produced in communicative contexts characterized by particularly conspicuous constraints” (Kruger and Van Rooy, 2016a:27), serves as an umbrella concept, bringing together two independent research areas—translation/interpreting and L2 research—in a unified framework (Kajzer-Wietrzny and Grabowski, 2021). Lanstyák and Heltai (2012) found that simplification and normalization are likely to be shared features of constrained communication, despite the differences between L2 production and translational communication. However, research in this area is still in its early stages, and more efforts are needed to understand the shared or non-shared features of these two constrained language varieties. Although good progress has been made in recent years, there is still much to learn about constrained language.

Research on constrained communication has mainly focused on translation and written L2 varieties, using various metrics such as multidimensional analysis (Kruger and Van Rooy, 2016a) and syntactic metrics (Ivaska and Bernardini, 2020; Ivaska et al., 2022) to detect commonalities and differences in different language varieties under different bilingual constraints. Few studies have focused on spoken constrained varieties, and those that have done so have mainly focused on features at the lexical level. For example, Kajzer-Wietrzny (2018) was the first to investigate spoken varieties using the framework of constrained languages. She showed that interpreters' use of the optional complementizer “that” resembled that of non-native speakers more than that of native speakers. Kajzer-Wietrzny and Ivaska (2020:175) found an “equalizing effect” in both spoken and written constrained varieties, where the interpretation of texts with typically literate features shifted toward the oral end of the continuum, while the interpretation of texts with typically oral features shifted toward the literate end (Shlesinger and Ordan, 2012:54). Kajzer-Wietrzny (2021) compared the types and frequency of cohesive devices used in constrained and non-constrained languages across both spoken and written modes and found that interpreters tended to use similar types of sentence transitions as translators, while L2 speakers often used enumerative types more frequently in spoken discourse.

To the best of our knowledge, this study may be one of the first to investigate simplification in non-native spoken constrained language varieties (interpreted and L2 speech) using syntactic complexity metrics. Additionally, a predominant number of studies using the theoretical framework of constrained communication have focused on bilingual production in European languages, while studies involving English and non-European languages are scarce. The present study focuses on Cantonese, which is a dialect widely spoken in Hong Kong and regarded as one of the seven major dialects in China (Chan and Li, 2000). Cantonese differs from English in vocabulary, semantics, and syntax, which may impose additional cognitive constraints on interpreters working between these two languages. Furthermore, both interpreters and L2 speakers in this study are working in their L2 (English), making these two spoken varieties more comparable due to the shared directionality (from L1 to L2).

2.4. Research questions

The main goal of this study is to address the following research questions (RQs):

(RQ1) How do spoken constrained languages (L1–L2 interpreted speech and original speech in L2) and non-constrained language (original speech in L1) compare in terms of the 14 syntactic complexity metrics? To what extent can simplification be identified as a shared characteristic in spoken constrained languages?
(RQ2) If simplification is confirmed as a shared characteristic in spoken constrained languages, what factors may explain the different syntactic complexity patterns between constrained and non-constrained languages in this study?
(RQ3) Will the two constrained languages (interpreted speech in L2 and original speech in L2) exhibit similar or different patterns in terms of syntactic complexity? What might be the possible reasons for any observed differences?

3. CORPORA AND METHODS

3.1. Corpora

Genre has been identified as an important factor influencing the syntactic complexity of translated texts (Al-Jabr, 2006; Liu and Afzaal, 2021; Xu and Li, 2021). However, due to the challenges of compiling spoken corpora (Castro, 2017; Campoy and Luzón, 2007), the present study limited its corpus material to the genre of political discourse. Hence, the Political Debate English Comparable Corpus (PDECC) was designed and constructed. It consists of three subcorpora: original speech in L1 (L1O), original speech in L2 (L2O), and interpreted speech in L2 (L2I). Each subcorpus includes 50 text files transcribed from speech and manually checked for unwanted textual noise. These texts were then tokenized and divided into chunks of approximately 2000 tokens ending at a sentence boundary, following the sampling

frame of the Brown Corpus (Francis and Kucera, 1964). This ensures that text length is not a confounding variable in the statistical analysis. The two constrained language varieties (L2O and L2I) were produced by L2 speakers with the same L1 background in Cantonese. As the source language of the interpreted speech is also Cantonese, the potential confounding influence of cross-language differences in the constrained language outputs due to different L1s can be ruled out.

To ensure the openness, availability, and comparability of data, debates in the House of Commons of the UK parliament were selected to compile the native English subcorpus (L1O), and simultaneous Cantonese-English interpretation of debates in the Legislative Council of Hong Kong (HK LegCo) was used to compile the interpreted English subcorpus (L2I). Consideration was also given to the similarity of topics, as topics have been identified as a factor affecting syntactic complexity (Yang et al., 2015). Both the UK parliament and the HK LegCo offer live broadcasts of all open meeting sessions, and the two subcorpora share similar settings, topics, and procedures, ensuring a high degree of comparability. The two subcorpora encompass a wide range of general political, social, and economic issues of public interest. The eight major topics include budget control/investment, social welfare, human rights, the economy, public/-government policy, housing policy, education and international relations. For the L2O subcorpus, two Hong Kong TV programs[1] featuring interviews on similar political and social topics were selected. These interviews, which involve conversations between hosts and interviewees (mostly politicians, businesspeople, and activists), are similar to the question-and-answer sessions in L1O and L2I. The speakers in the L2O corpus are non-native English as a Second Language (ESL) users whose L1 is Cantonese. The data for all three subcorpora were collected within the same time span, i.e., 2016–2020, and each subcorpus had a size of approximately 100,000 words, maximizing comparability in terms of setting, content, size, and time span.

To transcribe the videos into written texts, we used iFLYTEK's machine transcription software, which claims a high level of accuracy of over 98% on its official website.[2] These transcriptions were then manually proofread against the official verbatim reports of the debates, known as the UK and Hong Kong Hansards, available on their respective official websites. To focus on complexity, we manually removed features such as false starts, fillers, hesitations, mispronunciation, and use of the mother tongue that might affect accuracy and fluency. This cleaning of the corpus data was carried out to ensure that the L2SCA syntactic complexity software could be effectively applied to analyze the spoken corpora. To further validate the reliability of the L2SCA in processing spoken and interpreted texts, we manually annotated the production units identified by the L2SCA in 600 randomly selected language segments (approximately 200 sentences in each subcorpus).

The three authors of this study labeled the production unit and syntactic structure boundaries based on the detailed description provided in Lu (2010). Any disagreements between us were resolved through discussion. We then compared the correlation between the syntactic complexity measures calculated based on manual annotation and those computed by the L2SCA. The results showed that the majority of correlations between the syntactic complexity values calculated by annotators and the L2SCA were strong, ranging from 0.731 to 0.999 (complete correlation results can be found in Appendix A). This indicates that the results generated by the L2SCA in the current study are reliable.

Table 1 shows the breakdown of the statistics for the three subcorpora. The L2I corpus contains 104,077 words, the L2O corpus has 105,186 words, and the L1O corpus consists of 103,097 words.

3.2. Measures of syntactic complexity

To gain a more comprehensive understanding of syntactic complexity in constrained languages, this study used the L2SCA tool (Lu, 2009, 2010) to compute all 14 measures within 5 constructs. These 14 measures cover major constructs of syntactic complexity at the clausal (e.g., number of subordinated and coordinated phrases per T-unit), phrasal (e.g., number of complex nominals per T-unit), and overall (e.g., mean length of sentences (MLS)) complexity levels. Table 2 provides descriptions of the 14 measures, including their codes and definitions. For all 14 measures, a higher value indicates a higher level of syntactic complexity.

As mentioned earlier, the L2SCA is reliable when applied to texts containing grammatically complete sentences. The speeches selected for this study were of high quality, as they were delivered by native and advanced L2 speakers and professional interpreters. In addition, manual cleaning of unwanted textual noise such as pauses and fillers was carried out to ensure that the sentences in the corpora were as grammatically complete as possible. The 14 measures of syntactic complexity (organized into five subconstructs) were computed and then exported to a Microsoft Excel file for further statistical analysis. These 14 indices of syntactic complexity were computed for each of the three subcorpora on a single-text basis, and statistical comparisons and analysis were conducted using the R statistical package.

[1] The two TV programs are Talk the Walk and All about Money, provided by the HK International Business Channel.
[2] https://www.iflytek.com/news/2302.

Table 1
Descriptive statistics of the Political Debate English Comparable Corpus.

Sub-corpora	Text count	Overall size	Mean size	Std. deviation
L2I	50	104,077	2,082	32
L2O	50	105,186	2,104	30
L1O	50	103,097	2,062	32

Table 2
Syntactic Complexity Metrics Based on Lu (2010:479).

Measure	Code	Definitions
Type 1: Length of production unit		
Mean length of clause	MLC	# of words / # of clauses
Mean length of sentence	MLS	# of words / # of sentences
Mean length of T-unit	MLT	# of words / # of T-units
Type 2: Sentence complexity		
Sentence complexity ratio	C/S	# of clauses / # of sentences
Type 3: Subordination		
T-unit complexity ratio	C/T	# of clauses / # of T-units
Complex T-unit ratio	CT/T	# of complex T-units / # of T-units
Dependent clause ratio	DC/C	# of dependent clauses / # of clauses
Dependent clause per T-unit	DC/T	# of dependent clauses / # of T-units
Type 4: Coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
Sentence coordination ratio	T/S	# of T-units / # of sentences
Type 5: Particular structure		
Complex nominals per clause	CN/C	# of complex nominals / # of clauses
Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units

4. RESULTS

This section aims to answer the three RQs outlined in Section 2.4. The analysis begins with an exploratory overview of syntactic complexity metrics across the constrained and non-constrained languages (RQ1 and RQ2). We then examine the differences in syntactic complexity metrics by focusing on relevant metrics between the two constrained varieties (RQ3).

4.1. Patterns of syntactic complexity across the constrained and non-constrained spoken varieties

Table 3 presents the mean values and standard deviations for the 14 metrics of syntactic complexity across the three subcorpora. For RQ1, we sought to determine whether there is a significant difference in syntactic complexity between the three language varieties. As shown in Table 3, the mean values of 13 out of the 14 indices in the non-constrained language (L1O) are higher than those in the constrained languages (L2I and L2O). To determine whether the mean complexity values significantly differed between the three language varieties, we conducted a one-way ANOVA test, followed by a post hoc test with Bonferroni correction. We calculated the variances of the 14 syntactic complexity measures on a single-text basis and evaluated the significance of differences across the three subcorpora.

The one-way ANOVA tests indicated statistically significant differences ($p < .05$) in all 14 syntactic complexity measures. We then conducted a post hoc multiple comparison test to determine whether significant differences existed between any two of the three groups. To minimize the risk of making a Type I error in the three post hoc comparisons, we set the alpha value at 0.05 and adjusted the least significant difference (LSD) p value for pairwise comparison to

Table 3
Mean Values for 14 Indices of Syntactic Complexity.

Measure	Code	L2I Mean (SD)		L2O Mean (SD)		L1O Mean (SD)	
Type 1: Length of production unit							
Mean length of clause	MLC	10.383	(2.231)	8.495	(0.849)	10.073	(0.847)
Mean length of sentence	MLS	19.119	(4.751)	17.740	(2.356)	24.693	(2.234)
Mean length of T-unit	MLT	16.965	(3.821)	15.934	(1.849)	20.129	(1.814)
Type 2: Sentence complexity							
Sentence complexity ratio	C/S	1.849	(0.303)	2.096	(0.274)	2.465	(0.273)
Type 3: Subordination							
T-unit complexity ratio	C/T	1.639	(0.193)	1.883	(0.212)	2.007	(0.202)
Complex T-unit ratio	CT/T	0.411	(0.071)	0.486	(0.070)	0.596	(0.071)
Dependent clause ratio	DC/C	0.347	(0.044)	0.415	(0.047)	0.463	(0.046)
Dependent clause per T-unit	DC/T	0.575	(0.138)	0.789	(0.169)	0.937	(0.185)
Type 4: Coordination							
Coordinate phrases per clause	CP/C	0.185	(0.099)	0.126	(0.048)	0.195	(0.041)
Coordinate phrases per T-unit	CP/T	0.302	(0.155)	0.235	(0.082)	0.389	(0.084)
Sentence coordination ratio	T/S	1.124	(0.083)	1.112	((0.049)	1.228	(0.066)
Type 5: Particular structure							
Complex nominals per clause	CN/C	1.101	(0.327)	0.935	(0.147)	1.219	(0.151)
Complex nominals per T-unit	CN/T	1.809	(0.587)	1.757	(0.308)	2.435	(0.307)
Verb phrases per T-unit	VP/T	2.196	(0.271)	2.466	(0.310)	2.701	(0.255)

0.0167 (0.05/3). Table 4 presents the results of the post hoc Bonferroni tests between any two of the three groups. The differences in syntactic complexity between the non-constrained variety (L1O) and the two types of constrained varieties (L2O and L2I) were statistically significant, with L1O being more complex in all 14 measures than L2O and more complex than L2I in 12 measures. These results support our initial hypothesis that non-constrained spoken language is more complex in terms of syntactic complexity than constrained spoken varieties. Based on these findings, we conclude that the simplification hypothesis is supported in both L1–L2 interpreted speech and L2 original speech.

Table 4
Differences in Mean Complexity Values between L2I, L2O and L1O.

Measure	Code	L2I vs L1O	L2O vs L1O	L2I vs L2O
Type 1: Length of production unit				
Mean length of clause	MLC	> -	< *	> *
Mean length of sentence	MLS	< *	< *	> *
Mean length of T-unit	MLT	< *	< *	> -
Type 2: Sentence complexity				
Sentence complexity ratio	C/S	< *	< *	< *
Type 3: Subordination				
T-unit complexity ratio	C/T	< *	< *	< *
Complex T-unit ratio	CT/T	< *	< *	< *
Dependent clause ratio	DC/C	< *	< *	< *
Dependent clause per T-unit	DC/T	< *	< *	< *
Type 4: Coordination				
Coordinate phrases per clause	CP/C	< -	< *	> *
Coordinate phrases per T-unit	CP/T	< *	< *	> *
Sentence coordination ratio	T/S	< *	< *	> -
Type 5: Particular structure				
Complex nominals per clause	CN/C	< *	< *	> *
Complex nominals per T-unit	CN/T	< *	< *	> -
Verb phrases per T-unit	VP/T	< *	< *	< *

Note: * significant (<0.0167) after Bonferroni correction (p <.05/3).
- insignificant (>0.0167).
> indicates the former is higher than the latter.
< indicates the former is lower than the latter.

4.2. Patterns of syntactic complexity between the two constrained varieties

In the second part of the analysis, we compared the two constrained language varieties (i.e., L2O and L2I). The descriptive statistics in Table 3 show that L2O and L2I do not produce a consistent pattern in terms of syntactic complexity metrics: L2I has higher mean values in eight metrics, while L2O has higher mean values in the other six metrics. The one-way ANOVA post hoc multiple comparison tests revealed significant differences in the mean values of 11 metrics (excluding MLS, MLT, and CN/T) between L2I and L2O. In what follows, we discuss the 14 metrics by categorizing them into five syntactic complexity subconstructs: the length of the production unit, the amount of subordination, the amount of coordination, the degree of phrasal sophistication, and overall sentence complexity.

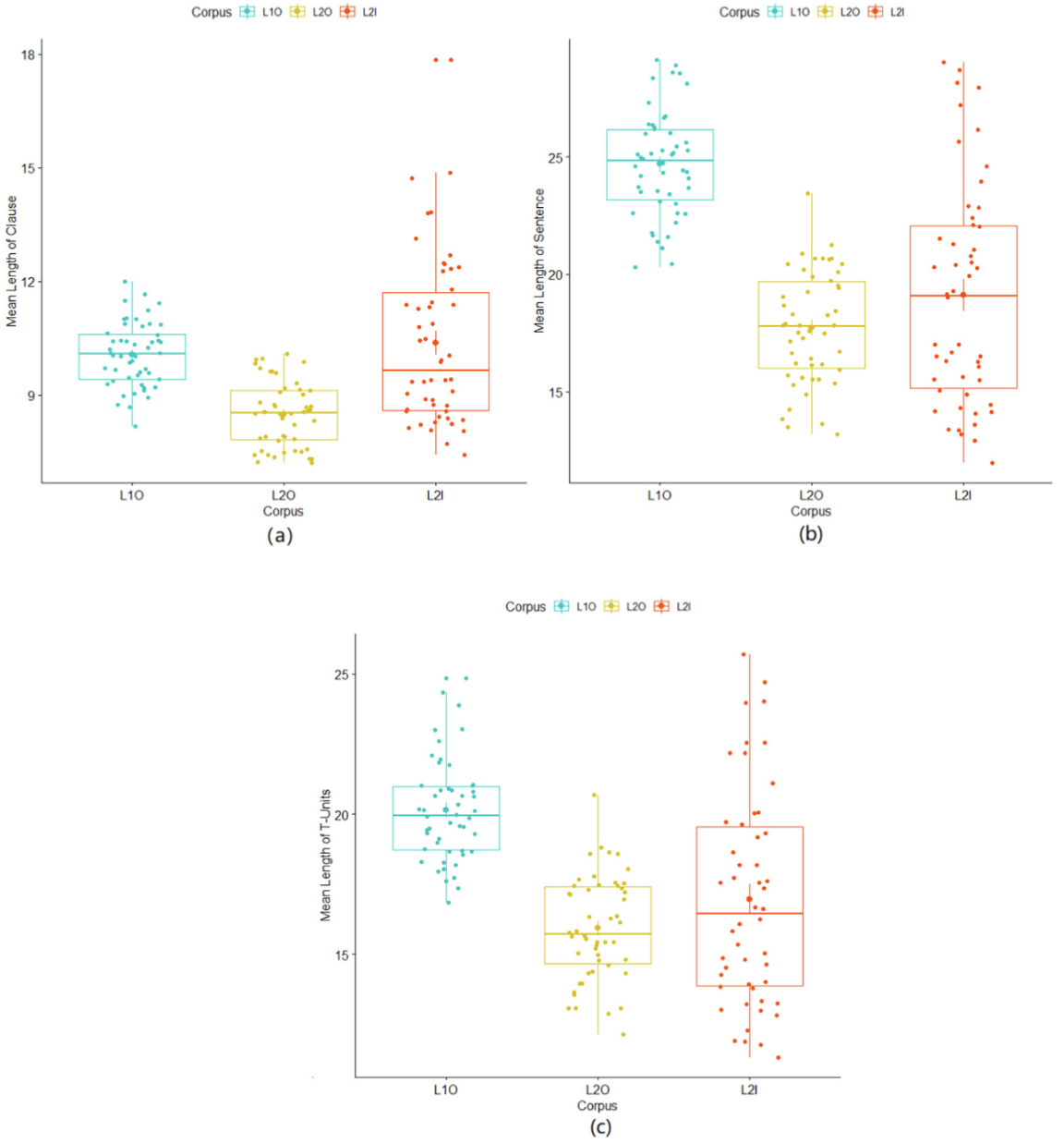

Fig. 1. **Syntactic complexity in the length of production unit:** (a) Mean length of clauses; (b) mean length of sentences; (c) mean length of T-units.

4.2.1. Length of the production unit

Among the three indices (mean length of clauses (MLC), MLS, and mean length of T-units (MLT)) in the construct of the length of the production unit, L2I has higher mean values than L2O, as shown in the boxplot figure (see Fig. 1). This result indicates that L1–L2 interpreted speech is syntactically more complex than L2 spontaneous speech. The post hoc tests with Bonferroni correction showed that L2I is significantly more complex than L2O in terms of the MLC and MLS but not in terms of the MLT. While L2I has a higher value than L1O in terms of the MLC, there is no significant difference between the two. Interestingly, the MLC is the only measure in which L2O is higher than L1O.

4.2.2. Amount of subordination

Fig. 2 illustrates the four subordination metrics among the three language varieties. As shown in Table 4 and Fig. 2, there are significant differences in all four subordination metrics among the three language varieties, with L1O being the

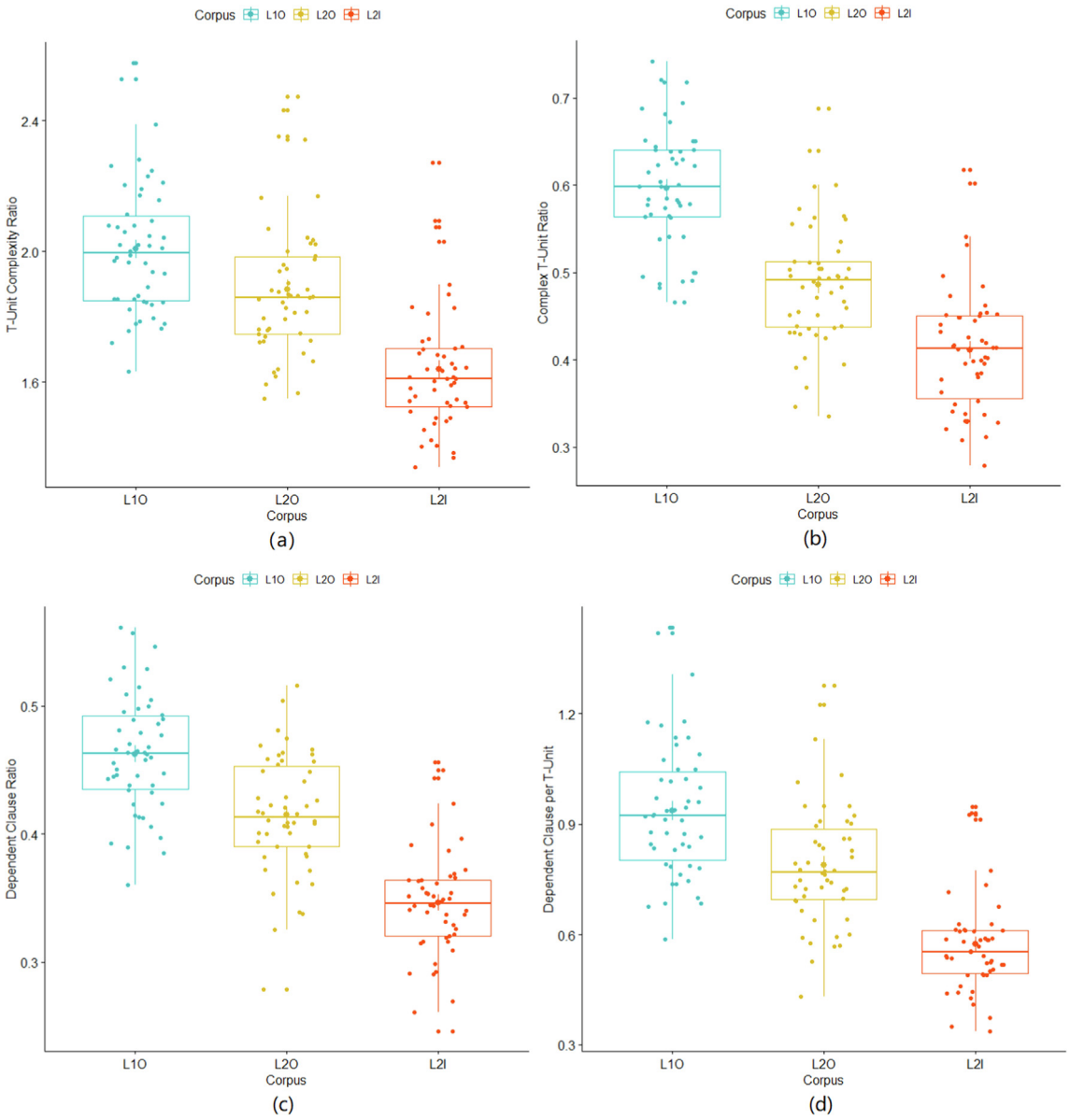

Fig. 2. **Syntactic complexity in the amount of subordination:** (a) The T-unit complexity ratio; (b) the complex T-unit ratio; (c) the dependent clause ratio; (d) dependent clauses per T-unit.

most complex, followed by L2O and L2I. Focusing on the two constrained language varieties, we see that L2O is more complex than L2I in all four metrics, indicating that L2O uses significantly more subordination than L2I. Overall, the interpreting variety is the least complex in terms of subordination use.

4.2.3. Amount of coordination

Fig. 3 illustrates the three coordination metrics among the three language varieties. As shown above, L1O is the most complex compared to the other two constrained language varieties. L1O is statistically more complex than L2O in all three metrics, and it is more complex than L2I in two metrics (coordinate phrases per T-unit (CP/T) and sentence coordination ratio (T/S)). In contrast to the trend seen in the subordination measures, L2I produces a greater degree of coordination than L2O in two metrics (coordinate phrases per clause (CP/C) and CP/T), while no significant difference was

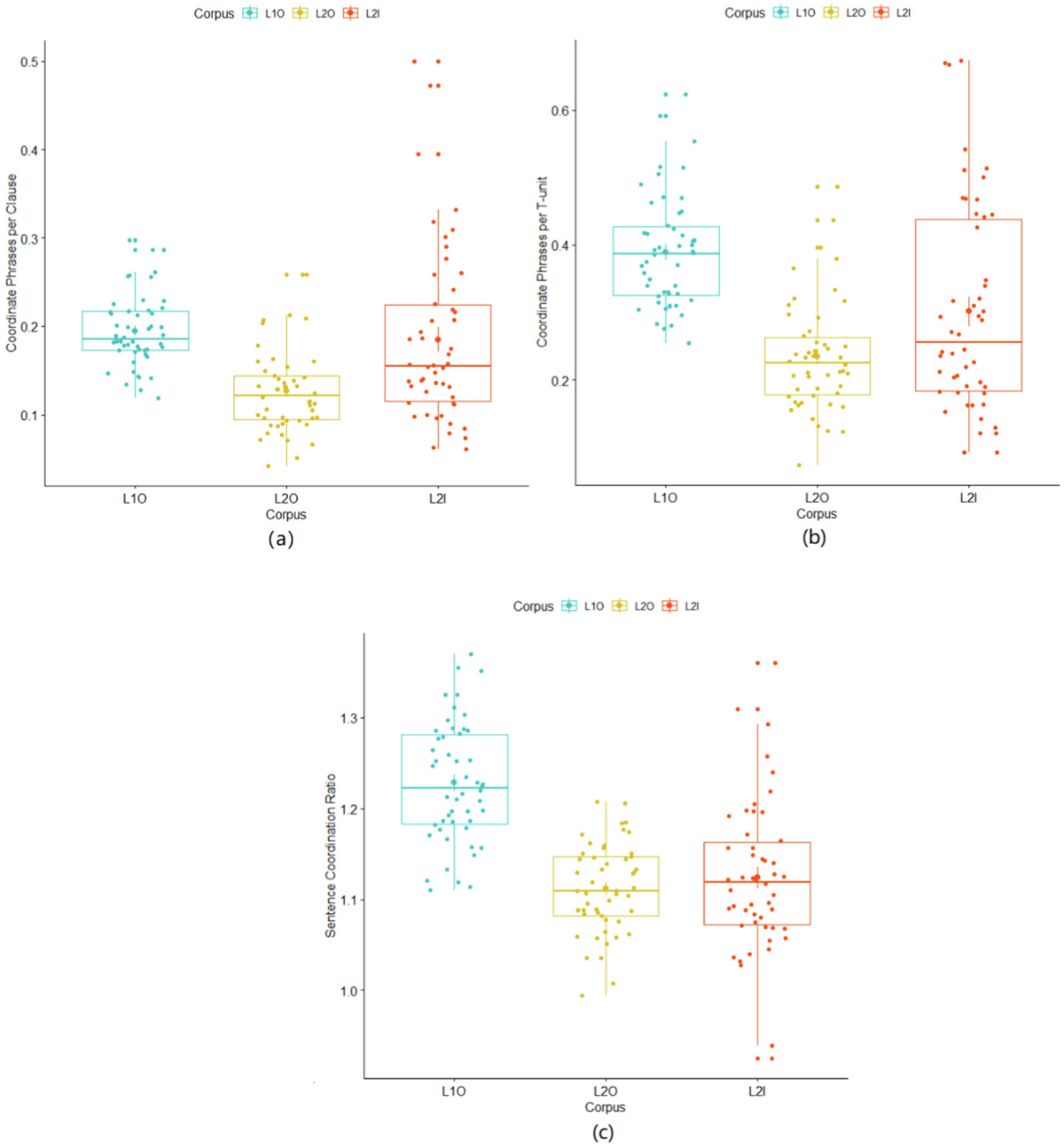

Fig. 3. **Syntactic complexity in the amount of coordination:** (a) Coordinate phrases per clause; (b) coordinate phrases per T-unit; (c) the sentence coordination ratio.

found in the T/S between the two groups. Overall, we see that L1–L2 interpreters tend to use more coordinate phrases than L2 speakers. However, both constrained language varieties are still not as complex as native speech in terms of coordination at both the clause and sentence levels.

4.2.4. Phrasal and overall sentence complexity

Fig. 4 illustrates the variation in phrasal and overall sentence complexity among the three varieties. As expected, L1O is the most complex in all four metrics compared to the other two varieties. In terms of the comparison of the two constrained languages, L2I uses significantly more complex nominals per clause, while L2O uses significantly more verb phrases per T-unit. No significant difference was found in the number of complex nominals per T-unit between the two constrained varieties. Regarding overall sentence complexity, L2O is significantly more complex than L2I, indicating that L2 speakers use more clauses per sentence than interpreters.

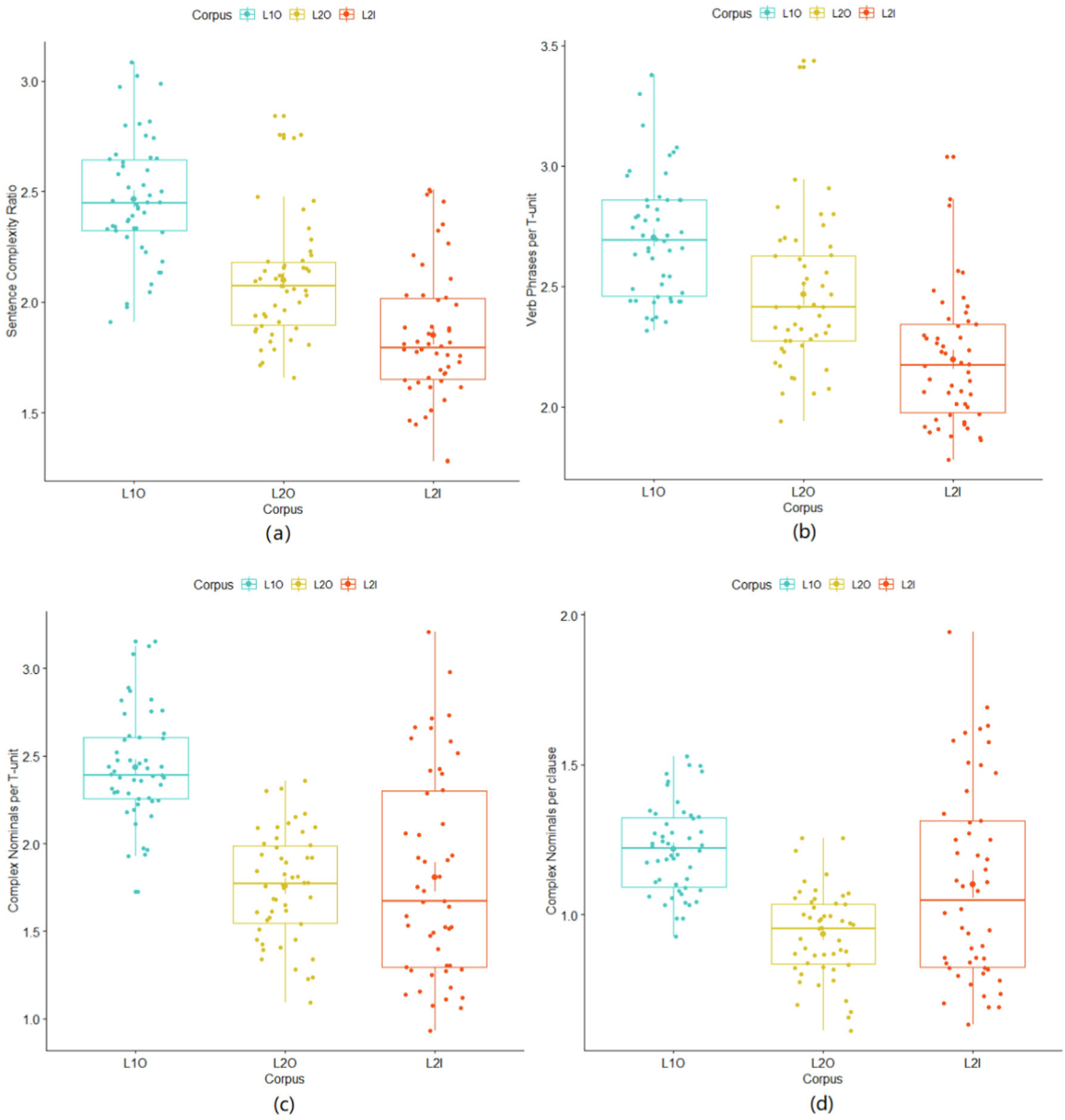

Fig. 4. **Syntactic complexity in phrasal and overall sentence complexity:** (a) The sentence complexity ratio; (b) verb phrases per T-unit; (c) complex nominals per T-unit; (d) complex nominals per clause.

5. DISCUSSION

The present study is one of the first to investigate simplification in spoken constrained language varieties using syntactic complexity metrics. A corpus of three subcorpora was compiled, including L1O, L2O, and L2I. The L2O and L2I subcorpora were produced by speakers with a Cantonese L1 background working in their L2 (English). The results showed that all 14 measures of syntactic complexity were significantly lower in both L2O and L2I conditions than in L1O conditions. When comparing L2O and L2I, significant differences were found in 11 of the 14 measures, with L2O being more complex in six measures and L2I in eight measures. These results support the hypothesis that constrained spoken languages are more simplified than non-constrained language. Additionally, they suggest that both L2 spoken language and L1–L2 interpreted language, which are believed to be under heavy cognitive load, operate at a lower level of syntactic complexity than native English speech.

The results of this study support the hypothesis that spoken languages under cognitive constraints, such as L2 speech and L1–L2 interpreted speech, tend to be less syntactically complex than non-constrained spoken languages. These findings are consistent with previous research indicating that interpreted English is simpler than L1 original English (Bernardini et al., 2016; Kajzer-Wietrzny, 2015; Sandrelli and Bendazzoli, 2005) and that L2 speech and interpreting are less complex than non-constrained varieties in terms of lexical use (Kajzer-Wietrzny and Ivaska, 2020). However, the relationship between syntactic and lexical complexity is not always straightforward, as other studies have found that translated Chinese can be simpler at the lexical level but not the syntactic level (Liu et al., 2022). Further research is needed to fully understand the complexity patterns of constrained languages in spoken modalities.

5.1. Effect of cognitive load in constrained languages

The current study shows that interpreting, as a constrained language variety, is characterized by syntactically simpler constructions than native speech. This phenomenon may be related to interpreters' techniques for coping with the high cognitive load of SI. The effort model proposed by Gile (2009) views SI as a process involving a series of operations on “successive speech segments” (Gile, 2009:168), including listening and analysis (Effort L), short-term memory (Effort M), speech production (Effort P), and coordination (Effort C). These efforts compete for limited processing resources and may be compromised due to the increased cognitive load from other efforts. Based on this model, Lv and Liang (2019) argue that interpreting output tends to have simpler linguistic features due to the high cognitive demands and challenges of the interpreting process. Our findings that SI exhibits lower levels of syntactic complexity compared to non-constrained texts support this argument. Clearly, SI, which is an extreme form of multitasking activity that involves intense cognitive load, can lead to more simplified language forms in the interpreter's output.

To reduce the cognitive load of SI, interpreters may use shorter chunks of the target language to maintain simultaneity with the source language, particularly when dealing with long and complex sentences (Kader and Seubert, 2014). This technique, known as “syntactic restructuring” (Riccardi, 2021: 378), involves breaking down long, syntactically complex sentences into shorter segments. By rendering the source language into shorter target language utterances, interpreters can alleviate the cognitive demands of SI and conserve their cognitive resources (Chernov, 2004; Cheung, 2012; Donato, 2003; Gile, 2009; Li, 2015; Shlesinger, 2003; Wu and Liao, 2018). In addition to the cognitive benefits of using shorter, simpler syntax, this technique may help interpreters speak more quickly and accurately under time pressure, as it requires less mental processing and reduces the risk of errors. When interpreting in an L2, interpreters often use “plain but clear” language and produce “short and simple sentences” (Wu and Liao, 2018:194) due to the added challenge of using the L2. This may contribute to the lower syntactic complexity level in L1–L2 interpreting compared to native speech.

For L1 output, native speakers have had extensive exposure to their language and have had the opportunity to learn and internalize the subtleties and nuances of its syntax. In other words, they have a greater range of syntactic structures at their disposal, as they have had more time to learn and internalize the syntax of their language. As a result, they can produce more varied and complex syntactic structures with greater ease. The finding of the current study that L1 output (L1O) is syntactically more complex than L2 output (L2O) is consistent with previous findings that the cognitive load of speaking in an L2 can reduce the language complexity in speech (Trebits, 2016). Adult learners often use simplification as a strategy to cope with the cognitive demands of learning a foreign language (Kortmann and Szmrecsanyi, 2009; Trudgill, 2001). Speech production can be more stressful than writing for L2 users due to time pressure and the lack of planning time, requiring effective coordination of conceptual, grammatical, and phonological encoding. As a result, L2 speakers may prioritize fluency over syntactic complexity (Larsen-Freeman, 2006, 2009; Ellis, 2008). This explains why L1 output (L1O) is syntactically the most complex among the three language varieties.

Simplification is often observed in constrained languages such as L2O and interpreted varieties compared to non-constrained language produced by native speakers (L1) (Kruger and Van Rooy, 2016a). Our study shows that both SI and L2 speech production are more cognitively demanding than L1 speech. The additional cognitive demands and time constraints of these constrained spoken varieties may inhibit the production of long and complex sentences, leading to a lower level of syntactic complexity.

5.2. Effect of the source-target language combination on syntactic complexity

In interpreting, syntactic complexity may also depend on the specific source-target language combination. While research on the syntactic complexity of interpreted speech in different language pairs is limited, studies on lexical simplification have shown that different language combinations may lead to mixed results in terms of simplification (Kajzer-Wietrzny, 2015). For example, Alonso Bacigalupe (2010) found that interpreters working between English and Spanish often produced syntactic structures that were parallel to the source texts to maintain the simultaneity of input and output under the linearity constraint.

The differences between typologically distinct languages, such as English and Chinese, can also affect the syntactic complexity of interpreted speech. These languages differ in terms of lexical features, syntactic rules and structures (Tsao, 1982) as well as word order, logical form, and the encoding of thematic and case relations (Setton, 1999). As a result, specific strategies may be required to overcome these differences. Dawrant (1996) found that interpreters working between Chinese and English often relied heavily on the strategy of linearity/segmentation to cope with the word-order differences between the two languages, potentially leading to lower levels of syntactic complexity in English renditions. It is very likely that the specific source-target language combination can be an important factor in determining the syntactic complexity of interpreted speech.

5.3. Coordination vs Subordination in constrained languages

To examine the relationship between sentence length in interpreting and L2 speech and coordination and subordination, we employed a multiple regression model. However, we first checked for multicollinearity using a stepwise variance inflation factor (VIF) selection procedure with the VIF function of the fmsb R package. We set the VIF threshold to 5 and found that two coordination variables, CP/C and CP/T, were highly correlated. Since a T-unit is a language unit that is closer to sentence length, we retained CP/T and removed CP/C. We also retained the syntactic coordination variable, T/S. Similarly, we found collinearity issues among the four subordination variables and removed the T-unit complexity ratio (C/T) and the complex T-unit ratio (CT/T), retaining the dependent clause ratio (DC/C) and dependent clauses per T-unit (DC/T). Our final model included four predictor variables (T/S, CP/T, DC/C, DC/T) and the outcome variable, the MLS.

The results of the regression analysis for L1–L2 interpretation are presented in Table 5. The model explained 79.1% of the variance in the MLS (R^2 = 0.791, adjusted R^2 = 0.772). We found that the MLS was influenced by both the CP/T and C/T, with the former having a stronger effect. In contrast, the T/S did not significantly impact the MLS. This result suggests that the longer sentence lengths in interpreted language are primarily due to the complexity of coordinate phrases rather than coordination at the sentence level. In the regression model for L2 speech (R^2 = 0.731, adjusted R^2 = 0.708), the MLS was affected by the C/T, followed by CP/T and T/S. Overall, these results indicate that L1–L2 interpreting is characterized by a higher use of coordinate phrases compared to L2 speech (as also shown in Fig. 3). This aligns with the findings of Wu et al. (2020) that interpreters tend to use more coordinate phrases to enhance clarity and make the source text more explicit. However, our results suggest that this explicitation is achieved primarily through the use of more coordinate phrases rather than through sentence subordination or coordination.

Table 5
Results of regressions on mean length of sentence.

	Mean length of sentence (L1-L2 Interpreting)			Mean length of sentence (L2 Speech)		
	β	*t*	*p*	*β*	*t*	*p*
T/S (Sentence coordination ratio)	0.0927	1.067	0.292	0.269	3.247	0.002
CP/T (Coordinate phrases per T-unit)	0.653	10.137	<0.001	0.281	3.887	<0.001
DC/C (Dependent clause ratio)	-0.070	-0.367	0.715	0.060	0.554	0.582
CT/T (Complex T-unit ratio)	0.355	2.983	0.005	0.449	4.000	<0.001

The regression model indicated that the sentence length of L2 speakers was largely influenced by their use of subordination, which was significantly higher than that of interpreters (as shown in Fig. 2). In addition, their sentence complexity ratio was significantly higher than that of interpreters (as shown in Fig. 4). This result aligns with previous research showing that L2 speakers tend to use more subordination as they become more proficient in their L2 (De Clercq and Housen, 2017; Tonkyn et al., 2012; Vercellotti, 2019). Given that the L2 speakers in our study were all advanced speakers, it is not surprising that they demonstrated a higher use of subordination.

On the other hand, interpreting is cognitively more demanding than L2 speech due to the added burden of mediation. As a result, we speculate that interpreters may use more coordinate phrases to manage this increased cognitive load rather than relying on sentence coordination or subordination. Both the interpreters and L2 speakers in our study were advanced language users, but they appeared to be operating under different cognitive constraints based on the results of the regression analysis.

6. CONCLUSION

The aim of the present study was to compare the syntactic complexity of spoken constrained varieties (L1–L2 interpreting and L2 speech) to that of non-constrained native speech. The results of the study provide strong evidence to support the hypothesis that constrained languages are less syntactically complex than non-constrained languages. Additionally, the two constrained varieties were found to be more similar to each other in terms of syntactic complexity measures than either was to the non-constrained spoken variety, which is consistent with previous research by Kruger and Van Rooy (2016a) on constrained written varieties (e.g., translation and L2 writing). However, the interpreting variety and L2 speech variety also differed from each other in terms of syntactic complexity patterns, suggesting that they may be subject to different cognitive constraints. Possible explanations for these findings are discussed in relation to theories of cognitive load in constrained languages, language combinations in interpreting, and the preference for subordinate and coordinate structures in constrained languages. It is hoped that this study will contribute to understanding constrained languages in the spoken modality by providing a clearer picture of the syntactic differences between constrained and non-constrained languages.

There are a few limitations to consider in the current study. One limitation is that the lower level of syntactic complexity observed in constrained languages may be specific to the language combination of Chinese (L1/source language) and English (L2/target language) used in this study. The current study did not consider the influence of the source language, as it followed the comparable corpus design introduced by Baker (1993). Future empirical studies could examine the variable of source language and its potential “shining-through” or interference effects (as described by Teich, 2003). In addition, the interpretations of the findings would benefit from being validated through a combination of process-oriented empirical research and corpus-based methods. Furthermore, the current study is limited to interpreting and L2 use in the Cantonese (Hong Kong) context. Future studies may involve different translation directions or language pairs to provide a more comprehensive understanding of syntactic complexity in constrained languages. Such an endeavor could yield a more nuanced understanding of the relationship between the source language and syntactic complexity in constrained languages.

Data availability

Corpus data concerning the study are publicly available on Open Science Framework (https://osf.io/gzjbw/)

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENTS

We are grateful to the editor and reviewers for their thoughtful suggestions and constructive feedback, which have significantly improved the quality and clarity of our work.

A. APPENDIX. CORRELATION OF COMPLEXITY SCORES: HUMAN ANNOTATION VS L2SCA

Measure	NE	IE	L2	Measure	NE	IE	L2
MLS	0.973	0.999	0.968	**DC/T**	0.937	0.914	0.893
MLT	0.915	0.970	0.969	**T/S**	0.853	0.899	0.954
MLC	0.775	0.895	0.731	**CT/T**	0.803	0.838	0.812
C/S	0.927	0.860	0.852	**CP/T**	0.904	0.926	0.915
VP/T	0.906	0.854	0.813	**CP/C**	0.917	0.959	0.909
C/T	0.899	0.798	0.805	**CN/T**	0.850	0.937	0.894
DC/C	0.885	0.907	0.809	**CN/C**	0.844	0.906	0.855

References

Al-Jabr, A.F., 2006. Effect of syntactic complexity on translating from/into English/Arabic. Babel 52 (3), 203–221.

Alonso Bacigalupe, L., 2010. Information processing during simultaneous interpretation: A three-tier approach. Perspectives 18 (1), 39–58.

Baker, M., 1993. Corpus linguistics and translation studies: Implications and applications. In: Baker, M., Francis, G., Tognini-Bonelli, E. (Eds.), Text and Technology: In Honour of John Sinclair. John Benjamins Publishing Company, pp. 233–252.

Baker, M., 1996. Corpus-based translation studies: The challenges that lie ahead. In: Somers, H. (Ed.), Terminology, LSP, and Translation: Studies in Language Engineering in Honour of Juan C. Sager. John Benjamins Publishing Company, pp. 175–186.

Bartłomiejczyk, M., 2006. Strategies of simultaneous interpreting and directionality. Interpreting 8 (2), 149–174.

Bernardini, S., Ferraresi, A., Miličević, M., 2016. From EPIC to EPTIC - Exploring simplification in interpreting and translation from an intermodal perspective. Target 28 (1), 61–86.

Bernardini, P., Granfeldt, J., 2019. On cross-linguistic variation and measures of linguistic complexity in learner texts: Italian, French and English. Int. J. Appl. Linguist. 29 (2), 211–232.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., 1999. Longman Grammar of Spoken and Written English. Longman, New York.

Biber, D., Gray, B., Poonpon, K., 2011. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? TESOL Q. 45 (1), 5–35.

Biber, D., Gray, B., Staples, S., 2016. Predicting patterns of grammatical complexity across language exam task types and proficiency levels. Appl. Linguis. 37 (5), 639–668.

Biber, D., Gray, B., Staples, S., Egbert, J., 2020. Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. J. Engl. Acad. Purp. 46, 100869.

Bulté, B., Housen, A., 2012. Defining and operationalising L2 complexity. In: Housen, A., Kuiken, F., Vedder, I. (Eds.), Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA. John Benjamins Publishing Company, pp. 23–46.

Campoy, M.C., Luzón, M.J., 2007. Spoken Corpora in Applied Linguistics, Vol. 51. Peter Lang.

Casal, J.E., Lee, J.J., 2019. Syntactic complexity and writing quality in assessed first-year L2 writing. J. Second. Lang. Writ. 44, 51–62.

Castro, T.C., 2017. Applications of simultaneous-interpreting corpora in terminology research. New Voice in Translation Studies 17, 31–51.

Chan, A.Y., Li, D.C., 2000. English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. Lang. Cult. Curric. 13 (1), 67–85.

Chernov, G.V., 2004. Inference and Anticipation in Simultaneous Interpreting: A Probability-prediction Model. John Benjamins Pub. Co., Amsterdam.

Cheung, A.-K.-F., 2012. Omission in simultaneous interpreting: Word order differences. Forum 10 (2), 19–33.

Christoffels, I.K., De Groot, A.M., 2004. Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. Biling. Lang. Congn. 7 (3), 227–240.

Dawrant, A., 1996. Word order in Chinese-English simultaneous interpretation: An initial exploration. Fu Jen University, Unpublished MA thesis.

Dayter, D., 2018. Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN). Forum 16 (2), 241–264.

De Clercq, B., Housen, A., 2017. A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity. Mod. Lang. J. 101 (2), 315–334.

Donato, V., 2003. Strategies adopted by student interpreters in SI: a comparison. between the English-Italian and the German-Italian language-pairs. The Interpreters' Newsletter 12, 101–134.

Ellis, R., 2008. The Study of Second Language Acquisition. Oxford University Press.
Ferraresi, A., Bernardini, S., Petrović, M., Lefer, M.-A., 2018. Simplified or not simplified? The different guises of mediated English at the European Parliament. Meta 63 (3), 717–738.
Francis, W.N., Kucera, H., 1964. Brown Corpus Manual., Department of Linguistics. Brown University, Providence, Rhode Island, US.
Gerver, D., 1976. Empirical studies of simultaneous interpretation: A review and a model. In: Brislin, R.W. (Ed.), Translation: Applications and Research. Gardner Press, New York, pp. 165–207.
Gile, D., 2009. Basic Concepts and Models for Interpreter and Translator Training Revised edition. John Benjamins Publishing Company.
Hild, A., 2011. Effects of linguistic complexity on expert processing during simultaneous interpreting. In: Alvstad, C., Hild, E., Tiselius, E. (Eds.), Methods and Strategies of Process Research. John Benjamins Publishing Company, pp. 249–267.
Hwang, H., Jung, H., Kim, H., 2020. Effects of written versus spoken production modalities on syntactic complexity measures in beginning-level child EFL learners. Mod. Lang. J. 104 (1), 267–283.
Hyönä, J., Tommola, J., Alaja, A.-M., 1995. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. Q. J. Exp. Psychol. 48 (3), 598–612.
Ivaska, I., Bernardini, S., 2020. Constrained language use in Finnish: a corpus-driven approach. Nordic Journal of Linguistics 43 (1), 33–57.
Ivaska, I., Ferraresi, A., Bernardini, S., 2022. Syntactic properties of constrained English: A corpus-driven approach. In: Granger, S., Lefer, M.-A. (Eds.), Extending the Scope of Corpus-Based Translation Studies. Bloomsbury Academic, pp. 133–157.
Jiang, X., Jiang, Y., 2020. Effect of dependency distance of source text on disfluencies in interpreting. Lingua 243, 102873.
Kader, S., Seubert, S., 2014. Anticipation, segmentation. . . stalling? How to teach interpreting strategies. In: Andres, D., Behr, M. (Eds.), To Know How to Suggest: Approaches to Teaching Conference Interpreting. Frank and Timme, pp. 125–144.
Kajzer-Wietrzny, M., 2012. Interpreting universals and interpreting style. Adam Mickiewicz University, Poznań, Poland, Unpublished PhD dissertation..
Kajzer-Wietrzny, M., 2015. Simplification in interpreting and translation. Across Languages and Cultures 16 (2), 233–255.
Kajzer-Wietrzny, M., 2018. Interpretese vs. non-native language use: The case of optional that. In: Russo, M., Bendazzoli, C., Defrancq, B. (Eds.), Making Way in Corpus-based Interpreting Studies. Springer, pp. 97–113.
Kajzer-Wietrzny, M., 2021. An intermodal approach to cohesion in constrained and unconstrained language. Target 34 (1), 130–162.
Kajzer-Wietrzny, M., Grabowski, Ł., 2021. Formulaicity in constrained communication. MonTI. Monografías de Traducción e Interpretación 13, 148–183.
Kajzer-Wietrzny, M., Ivaska, I., 2020. A multivariate approach to lexical diversity in constrained language. Across Languages and Cultures 21 (2), 169–194.
Khushik, G.A., Huhta, A., 2020. Investigating syntactic complexity in EFL learners' writing across common European framework of reference levels a1, a2, and b1. Appl. Linguis. 41 (4), 506–532.
Kortmann, B., Szmrecsanyi, B., 2009. World Englishes between simplification and complexification. In: Hoffmann, T., Siebers, L. (Eds.), World Englishes - Problems, Properties and Prospects: Selected Papers from the 13th IAWE Conference. Benjamins, pp. 265–286.
Kotze, H., 2022. Translation as constrained communication. In: Granger, S., Lefer, M.-.-A. (Eds.), Extending the Scope of Corpus-Based Translation Studies. Bloomsbury, London, pp. 67–97.
Kroll, J., De Groot, A., 2005. Simultaneous interpreting: a cognitive perspective. In: Kroll, J.F., De Groot, A.M.B. (Eds.), Handbook of Bilingualism. Oxford University Press, pp. 454–479.
Kruger, H., Van Rooy, B., 2016a. Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. English World-wide 37 (1), 26–57.
Kruger, H., Van Rooy, B., 2016b. Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. Lang. Sci. 56, 118–131.
Kuiken, F., Vedder, I., 2011. Task complexity and linguistic performance in L2 writing and speaking. In: Robinson, P. (Ed.), Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance. John Benjamins, pp. 91–104.
Kyle, K., Crossley, S.A., 2018. Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. Mod. Lang. J. 102 (2), 333–349.
Kyle, K., 2016. Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication. Unpublished doctoral dissertation. Georgia State University, Atlanta, GA.
Lambert, C., Kormos, J., 2014. Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. Appl. Linguis. 35 (5), 607–614.
Lanstyák, I., Heltai, P., 2012. Universals in language contact and translation. Across Language and Cultures 13 (1), 99–121.
Larsen-Freeman, D., 2006. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. Appl. Linguis. 26 (4), 590–619.
Larsen-Freeman, D., 2009. Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. Appl. Linguis. 30 (4), 579–589.
Laviosa, S., 1998a. Core patterns of lexical use in a comparable corpus of English narrative prose. Meta 43 (4), 557–570.

Laviosa, S., 1998b. The corpus-based approach: A new paradigm in translation studies. Meta 43 (4), 474–497.
Li, X., 2015. Putting interpreting strategies in their place: Justifications for teaching strategies in interpreter training. Babel 61 (2), 170–192.
Liang, J., Lv, Q., Liu, Y., 2019. Quantifying interpreting types: Language sequence mirrors cognitive load minimization in interpreting tasks. Front. Psychol. 10, 285.
Liu, K., Afzaal, M., 2021. Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification. PLoS One 16 (6), e0253454.
Liu, K., Liu, Z., Lei, L., 2022. Simplification in translated Chinese: An entropy-based approach. Lingua 275, 103364.
Lu, X., 2009. Automatic measurement of syntactic complexity in child language acquisition. Int. J. Corpus Linguis. 14 (1), 3–28.
Lu, X., 2010. Automatic analysis of syntactic complexity in second language writing. Int. J. Corpus Linguis. 15 (4), 474–496.
Lu, X., 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. TESOL Q. 45 (1), 36–62.
Lu, X., 2017. Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. Lang. Test. 34 (4), 493–511.
Lu, X., Ai, H., 2015. Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. J. Second. Lang. Writ. 29, 16–27.
Lv, Q., Liang, J., 2019. Is consecutive interpreting easier than simultaneous interpreting? – A corpus-based study of lexical simplification in interpretation. Perspectives 27 (1), 91–106.
McNamara, D.S., Graesser, A.C., McCarthy, P.M., Cai, Z., 2014. Automated Evaluation of Text and Discourse with Coh-Metrix. Cambridge University Press, Cambridge.
Norris, J.M., Ortega, L., 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. Canadian J. Appl. Linguis. 30 (4), 555–578.
Ortega, L., 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. Appl. Linguis. 24 (4), 492–518.
Pallotti, G., 2015. A simple view of linguistic complexity. Second. Lang. Res. 31 (1), 117–134.
Park, S., 2022. A corpus-based comparison of syntactic complexity in spoken and written learner language. Canadian J. Appl. Linguis. 25 (2), 47–70.
Riccardi, A., 2021. Strategies and capacity management in conference interpreting. In: Albl-mikasa, M., Tiselius, E. (Eds.), The Routledge Handbook of Conference Interpreting. Routledge, pp. 371–385.
Sandrelli, A., Bendazzoli, C., 2005. Lexical patterns in simultaneous interpreting: A preliminary investigation of EPIC (European Parliament Interpreting Corpus). Proceedings from the Corpus Linguistics Conference Series 1. University of Birmingham, Birmingham.
Seeber, K.G., 2011. Cognitive load in simultaneous interpreting: Existing theories—new models. Interpreting 13 (2), 176–204.
Seeber, K.G., Kerzel, D., 2012. Cognitive load in simultaneous interpreting: Model meets data. Int. J. Biling. 16 (2), 228–242.
Setton, R., 1999. A Cognitive-pragmatic Analysis of Simultaneous Interpretation. John Benjamins.
Shlesinger, M., 2003. Effects of presentation rate on working memory in simultaneous interpreting. Interpreters' Newsletter 12, 37–49.
Shlesinger, M., Ordan, N., 2012. More spoken or more translated?: Exploring a known unknown of simultaneous interpreting. Target 24 (1), 43–60.
Teich, E., 2003. Cross-linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts. Mouton de Gruyter, Berlin.
Tommola, J., Niemi, P., 1986. Mental load in simultaneous interpreting: An on-line pilot study. In: Evensen, L. (Ed.), Nordic Research in Text Linguistics and Discourse Analysis. Tapir, pp. 171–184.
Tonkyn, A., Housen, A., Kuiken, F., Vedder, I., 2012. Measuring and perceiving changes in oral complexity, accuracy and fluency. In: Housen, A., Kuiken, F., Vedder, I. (Eds.), Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA. John Benjamins, Amsterdam, pp. 221–245.
Trebits, A., 2016. Sources of individual differences in L2 narrative production: The contribution of input, processing, and output anxiety. Appl. Linguis. 37 (2), 155–174.
Trudgill, P., 2001. Contact and simplification: Historical baggage and directionality in linguistic change. Linguistic Typology 5 (2/3), 371–374.
Tsao, F., 1982. English and Chinese (Mandarin). Annu. Rev. Appl. Linguist. 3, 99–117.
Vercellotti, M.L., 2019. Finding variation: assessing the development of syntactic complexity in ESL Speech. Int. J. Appl. Linguist. 29 (2), 233–247.
Wu, Y., Liao, P., 2018. Re-conceptualising interpreting strategies for teaching interpretation into a B language. Interpreter Translator Trainer 12 (2), 188–206.
Wu, X., Mauranen, A., Lei, L., 2020. Syntactic complexity in English as a lingua franca academic writing. J. Engl. Acad. Purp. 43, 100798.
Xu, J., Li, J., 2021. A syntactic complexity analysis of translational English across genres. Across Languages Cultures 22 (2), 214–232.
Yang, W., Lu, X., Weigle, S.C., 2015. Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. J. Second. Lang. Writ. 28, 53–67.