Available online at www.sciencedirect.com

ScienceDirect

Lingua 294 (2023) 103607

Lingua

www.elsevier.com/locate/lingua

Syntactic simplification in interpreted English: Dependency distance and direction measures

Han Xu, Kanglong Liu *

Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China

Received 17 May 2023; revised 3 September 2023; accepted in revised form 5 September 2023;

Abstract

This study investigates the simplification hypothesis in interpreting, as well as its cognitive implications, by examining features of syntactic dependency in three language varieties: English speech simultaneously interpreted from Chinese, original English speech produced by native speakers (L1 speech), and original English speech produced by non-native speakers (L2 speech). Two measures of the dependency relation, namely dependency distance and dependency direction, are employed to explore the distinction among the three language varieties in terms of their syntactic complexity, amount of cognitive demand, and the typological property of word order. The findings reveal that interpreted speech has the lowest mean dependency distance (MDD), followed by L2 speech and L1 speech, which indicates that interpreted English speech is syntactically more simplified than original English speech. The lowest MDD in interpreted speech is associated with the high cognitive demand in simultaneous interpreting, suggesting that increased cognitive demand in language processing is likely to lead to simplification of the syntactic structure of the linguistic output. Furthermore, dependency direction analysis of the three language varieties indicates that interpreted English tends to be more head-final than L1 English speech, confirming a typological word order distinction between translational and original language.

Keywords: Dependency distance; Dependency direction; Simplification; Interpreting; Cognitive load

1. INTRODUCTION

Corpus-based interpreting studies have experienced great popularity over the past two decades as providing the field with a data-driven methodology that allows for systematic analysis of large amounts of machine-readable data obtained from authentic interpreting activity (Shlesinger, 2000; Shlesinger and Ordan, 2012). The interest in this approach lies in the identification of distinctive properties distinguishing interpreted speech from other language varieties (Bendazzoli, 2018, Shlesinger, 1998). Empirical research, by comparing interpreted speech with other language varieties, including original non-interpreted speech (Bendazzoli and Sandrelli, 2005) or translated texts (Bernardini

* Corresponding author.
E-mail addresses: hanxu@polyu.edu.hk (H. Xu), kl.liu@polyu.edu.hk (K. Liu).

https://doi.org/10.1016/j.lingua.2023.103607
0016-7037/

et al., 2016; Shlesinger and Ordan, 2012; Xu and Li, 2022), has identified several such properties, such as simplification (Bernardini et al, 2016; Liu and Afzaal, 2021; Liu et al., 2023), interference (Ma and Cheung, 2020) and explicitation (Gumul, 2021). Despite these findings, scholars maintain differing perspectives on the universality of the properties identified (Baker, 1993; Shlesinger, 1998). While some of these have been suggested as dominant patterns, such assertions remain disputed due to inconsistent results found in prior research (Chesterman, 2010; Kotze, 2020; Pym, 2008). Nevertheless, the application of corpus-based methodology to the study of interpreting activity serves to enhance our understanding of its linguistic patterns and regularities.

Of the various properties identified, the phenomenon of simplification, which refers to interpreters' tendency to use simpler language than the source text in their renditions, has received significant attention (Bernardini et al., 2016; Kajzer-Wietrzny, 2012; Liu et al., 2023; Lv and Liang, 2019; Xu and Li, 2022). As a distinctive linguistic feature, simplification was first examined in corpus-based translation studies. Many investigations have adduced significant evidence supporting its existence in translated texts, as shown in work by Kruger and Van Rooy (2016) and Laviosa (1998). However, when it comes to interpreting, research on the use of simplification has not yet reached definitive conclusions, and most studies have focused on simplification only at the lexical level (e.g., Bernardini et al., 2016; Kajzer-Wietrzny, 2012; Lv and Liang, 2019), which raises the question of whether and how this phenomenon manifests itself at other levels of language.

From the viewpoint of translation universals, Chesterman (2010, p. 43) posits a "cognitive cause" for the identified translation properties, thus providing a potential explanation for their presence. In a similar vein, De Sutter and Lefer (2020) contend that the question of how cognitive mechanisms shape translation remains a crucial, yet unanswered, issue in corpus-based translation studies. Since interpreted speech can be considered a final product of cognitive processes, it contains features that may be traced back to reveal the cognitive effort required to produce such an output (Shlesinger, 2000). However, few studies have explored the relationship between the properties of interpreted speech and their underlying cognitive mechanisms. Corpus-based investigations of interpreted speech and studies examining the cognitive aspects of interpreting have remained largely separate research streams. Most studies employing a corpus-based approach to identifying linguistic regularities have not explicitly investigated the cognitive aspects of interpreting or used them to explain the results. This research trend may be influenced by the primarily descriptive focus of corpus linguistics, which is to describe rather than explain observed phenomena (Anderman and Rogers, 2008; Neumann and Serbina, 2020).

Given the shaping effect of human cognition on linguistic structures (Bialystok and Craik, 2010; Levon and Buchstaller, 2015), it is reasonable to suggest that the linguistic representation of the properties of interpreted speech will reflect the cognitive processes involved in interpreting. To investigate the phenomenon of simplification in interpreting and its cognitive implications, this study used the Political Debate English Comparable Corpus (PEDEC), consisting of three sub-corpora of original English speech produced by native speakers (L1 speech), original English speech produced by speakers who use English as a second language (L2 speech) and speech interpreted into English from Chinese (interpreted speech). The study characterized syntactic dependency relations in these three language varieties (Hudson, 2010; Liu, 2008) in order to uncover the dynamic interplay between the properties of interpreted speech, its textual operationalization, and its cognitive costs. This investigation aims to enhance our understanding of the cognitive processes involved in interpreting and how they shape syntactic structures in interpreted speech. Following the introduction, Section 2 reviews relevant literature on simplification and cognition in interpreting, and introduces the concept of dependency distance, the key measurement of syntactic complexity used in this study. Section 3 provides an overview of the corpus and its process of compilation, followed by a description of the process of data analysis. Section 4 reports on the results of the data analysis, and Section 5 discusses their cognitive implications. Finally, Section 6 concludes the study by summarizing the main findings and suggesting potential directions for future research.

2. LITERATURE REVIEW

2.1. Simplification in interpreting

Over the years, researchers have increasingly investigated whether simplification can be considered a distinguishing feature of interpreting. Simplification is primarily operationalized in these studies at the lexical level, through the examination of such parameters as standardized type-token ratio, lexical density, list head coverage, core vocabulary coverage, which indicate the level of lexical variety, informativeness, and repetition (Bendazzoli and Sandrelli, 2005; Ferraresi et al., 2018; Kajzer-Wietrzny, 2013). This approach has its origin in Laviosa's (1998) seminal work, which probes lexical simplification in translation from a monolingual comparative perspective by comparing translated texts with original non-translated texts. More recently, researchers have taken an intermodal perspective by examining simplification across varying types of mediation, such as between translated and interpreted texts (Bernardini et al., 2016;

Ferraresi et al., 2018; Liu et al., 2023; Xu and Li, 2022) or among different types of interpreting such as consecutive v. simultaneous interpreting (Lv and Liang, 2019).

Most existing studies suggest an overall effect of lexical simplification resulting from the type of mediation process, such as translation v. interpreting, with such an effect being observed more prominently in interpreting than in translation (Bernardini et al., 2016; Ferraresi et al., 2018; Xu and Li, 2022). However, studying specific aspects of simplification, as represented by various lexical parameters, do not always yield coherent results, something which seems to challenge the simplification hypothesis. For example, in their study based on an intermodal corpus consisting of original legislative proceedings in Hong Kong and their translated and interpreted texts from Cantonese into English, Xu and Li (2022), found that interpreted speech is less lexically varied and more repetitive than both original speech and translated texts, showing a strong simplification effect. However, they also found that the interpreted speech is more informative than original speech, and is characterised by a higher degree of lexical density. This trend for interpreted speech to be more informative than the original speech was also found by Russo and colleagues (2006) and Kajzer-Wietrzny (2012). However, in another larger-scale study of simplification based on an intermodal corpus of both interpreted and translated texts, Bernardini and colleagues (2016) found that all the lexical measures, including lexical density, coherently indicate a strong simplification effect in interpreting when interpreted speech was compared to both original speech and translation. They further revealed that depending on the language combination, there may be different patterns of simplification. Specifically, the study found that interpreted Italian speech exhibited a significantly lower lexical density and shorter mean sentence length than its translated versions, while interpreted English speech was characterized by more frequent use of words, resulting in greater repetition compared to the original speech.

The varying results obtained from previous studies suggest that simplification should not be viewed as an overarching pattern but rather as a dynamic feature of interpreting that is manifested differently depending on the measures and context involved. Yet, an exclusive focus on lexical simplification may not fully capture its complexity and dynamic nature, and so it is necessary to investigate the simplification hypothesis using diverse approaches and valid measures that allow for its operationalization at different linguistic levels. Liu et al. (2023) recently examined 14 syntactic parameters to assess the level of syntactic complexity in three language varieties: interpreted speech, L1 speech, and L2 speech. Their findings largely support the simplification hypothesis, demonstrating that “constrained language” (i.e., interpreted speech and L2 speech) is syntactically simpler than unconstrained language (L1 speech), with interpreted speech being the most simplified. This innovative study expands the understanding of simplification beyond the lexical level and introduces L2 speech as a new comparable language variety. It provides valuable insights into interpreting and points to a promising direction for future research on simplification in interpreting.

2.2. Interpreting as a dynamic cognitive process

It is noteworthy that the majority of studies on simplification in interpreting focus solely on surface-level features and offer limited understanding of the underlying motivations behind the observed linguistic patterning. Yet, understanding what happens in the interpreter’s mind during interpreting is crucial. Research has long been invested in uncovering the cognitive mechanisms at play during interpreting, with the aim of conceptualizing the interpreting process, identifying specific tasks that may pose challenges for interpreters, and comprehending the various strategies utilized by interpreters to handle processing demands (Rennert, 2008; Shlesinger, 2000; Seeber and Kerzel, 2012). Although the cognitive aspects of interpreting are frequently discussed, the term “cognitive load” lacks a clear definition and is often used interchangeably with other terms such as cognitive demand, mental effort, and memory load (Chen, 2017). Seeber (2011) is one of the few to attempt to provide a working definition, stating that cognitive load is “the amount of capacity the performance of a cognitive task occupies in an inherently capacity-limited system”. Chen (2017, p. 643) argued that in interpreting, cognitive load is a multidimensional construct, representing an interaction between two groups of variables: task and environmental characteristics, and interpreter characteristics. In line with Seeber and Chen's definitions, research in this area aims to explore how an interpreter's cognitive capacity varies with changes in specific tasks or environmental factors, as well as his/her own characteristics. Existing evidence suggests that variations in cognitive load may arise during an interpreter’s work due to differences in syntactic constructions (Seeber and Kerzel, 2012), types of interpreting (Liang et al., 2017), and directionality (Rinne et al., 2000).

As a theoretical notion, the quantification of cognitive load for interpreters cannot be calculated directly but requires appropriate operationalization (Neumann and Serbina, 2020; Stefanowitsch and Flach, 2017). To describe and predict cognitive load in interpreting, two analytical frameworks, namely Gile’s Effort Model (2009) and Seeber's Cognitive Load Model (2011), have been widely adopted. While both models were originally designed for simultaneous interpreting, they can be adapted to account for cognitive activity in other types of interpreting as well (Bóna and Bakti, 2020; Liang et al., 2017). Other researchers have attempted to measure cognitive load in interpreting through the use of physio-psychological indicators such as eye movement, keystroke logging, and pen recording (Chen, 2017; Seeber,

2013). For example, Seeber and Kerzel (2012) found that different syntactic structures impose varying cognitive load on interpreters, as evidenced by the extent of pupil dilation. Alternatively, cognitive load can be assessed by evaluating the interpreter's level of performance (Bóna and Bakti, 2020; Swabey et al., 2016) or through the use of interpreter's retrospective self-report, or a combination of the two (Gumul, 2021). In a study investigating the impact of the complexity of a speech task on the interpreter's performance, Bóna and Bakti, (2020), in order to evaluate the interpreter's level of fluency used the measure of temporal characteristics and disfluency patterns as parameters of cognitive load. The study found that cross-lingual language production in the case of consecutive interpreting and sight translation poses a higher cognitive demand on interpreters compared to monolingual speech production.

Many studies examining the cognitive aspects of interpreting have primarily focused on simultaneous interpreting (Gile, 2009; Gumul, 2021; Liu et al., 2004; Seeber, 2011, 2013). This may be because simultaneous interpreting involves a complex real-time combination of a language comprehension task and a language production task, making it more cognitively demanding than other types of interpreting (Seeber, 2011). As a result, the inherent multitasking nature of simultaneous interpreting provides an ideal context for interpreting researchers to explore the cognitive mechanisms underlying interpreting activity and for cognitive science researchers to test their theories and models of cognition (Christoffels and De Groot, 2004, 2009). Following this research trend, the present study has adopted the concept of dependency distance, a syntactic measure of cognitive load, to examine the cognitive demand involved in simultaneous interpreting compared to other language varieties. The next sub-section provides an introduction to dependency distance and its application in investigating syntactic complexity and the associated cognitive demand in language processing.

2.3. Measuring dependency relations

During the process of human parsing, which involves the syntactic analysis of a sentence, a word remains in working memory until it forms a dependency relation with another word (Liu et al., 2009, p. 165). The dependency relations between two words that depend on or syntactically govern one another form the syntactic structure of a sentence (Liu, 2008; Hudson, 2010). Dependency distance (DD) is a measure that describes the separation between a word and its governor (also known as parent or linguistic head) or dependent (Liu et al., 2017), and it can be quantified by the number of intervening words between them (Hudson, 2010). As human working memory is limited, the longer the dependency distance between two words, the greater the memory cost of maintaining the two syntactically related words without losing information (Gibson, 1998; Grodner and Gibson, 2005). Due to the way it reflects different kinds of syntactic construction, DD is a crucial measure used to assess syntactic complexity and the cognitive cost of processing dependency in various research contexts.

Previous studies have employed dependency distance (DD) as a measure for typological analyses of different languages (Futrell et al., 2015; Liu, 2008, 2009; Liu et al., 2017). Liu's (2008) research, which calculated the average DD of 20 languages, discovered a dependency distance threshold for most human languages. Due to the limited capacity of human working memory and the constraints imposed by grammar, natural languages for the sake of efficiency typically reduce their DD, resulting in an average DD of no more than three words. This tendency has been further supported by corpus-based statistical analyses that compared the DD of human-produced sentences with those of sentences with computer-generated randomised word orders (Futrell et al., 2015; Liu et al., 2017; Liu, 2008). In their investigation of a corpus of 37 diverse languages, Futrell and colleagues (2015) found that the overall DD for all languages investigated was shorter than the random baseline, confirming that dependency distance minimisation (DDM) is a universal characteristic of human languages. Futrell and colleagues (2015: 10336) argued that DDM "is well motivated because it allows for more efficient parsing and generation of natural language". In a separate investigation analysing the diachronic variation in dependency distance (DD) of State of the Union addresses delivered by American presidents over a period of two hundred years, Lei and Wen (2020) observed that over a long temporal span DD tended to minimise.

More recently, there has been an emerging body of research utilizing DD to investigate various cross-linguistic processing issues and elucidate their related cognitive processes (Fan and Jiang, 2019; Liang et al., 2017; Liang and Sang, 2022; Shen et al., 2023; Wang and Liu, 2013; Yan and Liang, 2022). Focusing on translation and interpreting, these studies demonstrated the great potential of DD as a tool to explore distinctive syntactic patterns in translational language (Fan and Jiang, 2019), distinguish different interpreting types (Liang et al., 2017), predict the difficulty of different interpreting tasks (Jiang and Jiang, 2020) and the associated level of disfluency in interpreting output (Shen et al., 2023), and measure the impact of foreign language learning anxiety on learning interpreting skills (Yan and Liang, 2022). Based on a dependency-annotated treebank that consists of translated English texts, native English texts, and native Chinese texts, Fan and Jiang (2019) found that the mean dependency distance (MDD) of the translated English texts is much higher than that of native English texts, indicating a higher cognitive demand in producing translations. Similarly, Liang and Sang (2022) found a higher DD in English abstracts translated from Chinese compared to abstracts

originally written in English, indicating greater cognitive demand in producing a translated work. The longer DD in the translated sentences was attributed to both the higher cognitive cost of translation and the influence of the source language. Specifically, Chinese has a longer DD than English, resulting in a "shining through" effect in the English texts (Teich, 2003).

However, research on interpreting shows a somewhat different result (Liang et al., 2017; Yan and Liang, 2022). Based on a dependency-annotated treebank of authentic interpreted speeches, Liang et al. (2017) compared the DD of three types of interpreting, that is consecutive interpreting (CI), simultaneous interpreting (SI) and read-out translated speech (TR), in order to explore their respective cognitive demands. They found that the three types of interpreting exhibit different DD, with TR having the largest and CI the smallest DD, regardless of the source speech text size, language type, or the interpreter's individual style. A shorter DD in interpreted speech (CI and SI) does not necessarily indicate a lower cognitive demand in interpreting. Instead, Liang and colleagues argued that interpreters need to work with a heavier memory burden, so they have to reduce DD to relieve the cognitive load, which complies with the universal language property of DDM. In the same vein, Yan and Liang (2022) found that when students are under greater foreign language anxiety, the DD of their consecutive interpreting is shorter, indicating a higher cognitive load in the process of performing interpreting tasks.

The studies mentioned above provide robust evidence for the potential use of dependency relations as a crucial metric for differentiating between various different languages and language varieties based on their dependency structures. The present study seeks to examine the simplification hypothesis in interpreting and its cognitive implications by analysing the dependency relations of simultaneous interpreting and two other language varieties, namely L1 speech and L2 speech. The study is aimed at addressing the following research questions:

RQ1: To what extent do the dependency relations of interpreted speech differ from those of L1 and L2 speech?

RQ2: How do the observed syntactic variations among the three language varieties relate to their respective cognitive demands in terms of processing and production?

3. METHODOLOGY

3.1. Corpus compilation

To investigate the proposed research questions, the present study constructed a Political Debate English Comparable Corpus (PEDCC) consisting of three language varieties: L1 English speech (L1O), L2 English speech (L2O), and speech simultaneously interpreted from Cantonese into English (L2I).[1] The L1O sub-corpus was compiled using debates from the House of Commons in the British Parliament, ensuring that all speech was produced by native English speakers (L1 speakers). The L2O sub-corpus was compiled using interviews from two popular Hong Kong TV programs, All About Money and Talk And Walk. All the interviews were conducted in English by native Cantonese speakers (L2 speakers), covering political, financial, and social topics. The L2I sub-corpus was compiled using renditions of the debates from the Legislative Council of Hong Kong (HK LegCo). The HK LegCo debates are conducted in Cantonese and are simultaneously interpreted into English by professional interpreters who are native Cantonese speakers.

The PEDCC is composed of three sub-corpora, each comprising 50 texts, totalling approximately 300,000 words in the entire corpus. All debates and interviews included in the PEDCC are publicly accessible. The texts were carefully selected from the same period spanning 2016 to 2020, with each sub-corpus containing texts discussing similar topics in a question-and-answer format. The selection process was designed to ensure comparability among the three sub-corpora in terms of size, time, genre, language type, and format. This approach was adopted to mitigate potential confounding factors that may impact the understanding of simplification in interpreting (Bernardini et al., 2016; Kajzer-Wietzny, 2013; Xu and Li, 2022). The selected texts were transcribed using iFLYTEK, an automatic transcription software, and then manually checked for accuracy. For L1O and L2O, the transcripts were cross-checked with the verbatim reports of the debates published online by the UK and Hong Kong Hansards.[2] To ensure accurate syntactic analysis, distracting speech features such as discourse markers such as hesitations and fillers, mispronunciations, and code switching were manually removed. Table 1 provides an overview of the three sub-corpora.

[1] The L1O and L2I are existing corpora that were initially compiled in Xu and Li (2022). The L2O is a new sub-corpus added to PEDCC.

[2] See https://hansard.parliament.uk.

Table 1
An overview of the Political Debate English Comparable Corpus (PDECC).

Sub-corpus	Texts count	Token no.	Source	Producer	Language
L1O	50	104,077	UK Parliament debates	L1 speakers	English
L2O	50	105,186	HK TV interviews	L2 speakers	English
L2I	50	103,097	HK LegCo debates	L2 speakers	English

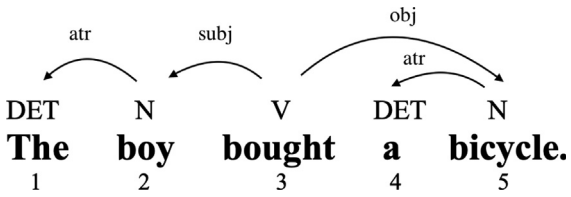

Fig. 1. Dependency relations of sample sentence “The boy bought a bicycle.”.

3.2. Calculating dependency relations

In this study, we utilized mean dependency distance (MDD) and dependency direction as measures of dependency relations to investigate simplification in interpretation and its cognitive implications. The syntactic structure of a sentence involves dependencies between individual words, as discussed by Hudson (2010) and Liu (2009). The core properties of syntactic dependency relations can be summarized in the following three principles: “1) it is a binary relation between two linguistic units; 2) it is usually asymmetrical, with one of the two units acting as the governor and the other as dependent; 3) it is classified in terms of a range of generally grammatical relations, as shown conventionally by a label on top of the arc linking the two units” (Liu, 2008, pp. 162-163). Based on these three principles, a directed acyclic graph that visually represents the dependency relations of a sentence can be constructed, as exemplified in Fig. 1.

In this example, all the words in the sentence are connected by dependency relations, which consist of a *governor* and a *dependent*, with a labelled arc extending from the *governor* to the *dependent*. For instance, the determiner *the* depends on the noun *boy* it modifies, and the subject *boy* depends on the verb *bought* that follows it. Thus, the noun forms dependency relations with the determiner and the verb, which serve as governor and dependent respectively. The position of each word in the sentence is marked by a number. The current method for calculating DD, introduced by Liu, Hudson, and Feng (2009, pp. 166-167), involves subtracting the numeric position of the dependent from that of the governor. For example, the DD between “boy” and “bought” is calculated as 3–2 = 1. If the numeric position of the dependent is higher than that of the governor, the DD can be a negative value, indicating a head-initial relationship. Conversely, if the numeric position of the dependent is lower than that of the governor, it represents a head-final relationship and yields a positive value of DD. However, the absolute value is used to measure the DD between two words. If a sentence contains only one word, the DD is zero, as a single word cannot form a dependency relation on its own. The MDD of a sentence can be obtained using the formula below:

$$MDD(sentence) = \frac{1}{n-1}\sum_{i=1}^{n-1}|DD_i|$$

In this formula, n represents the number of words in the sentence, and DDi represents the dependency distance of the i-th syntactic link in the sentence. Typically, in a sentence, there is one root verb that does not have a governor, like the word “bought” in example 1. The dependency distance of this root verb is considered as zero. Using this formula, the MDD of the sentence in Fig. 1 can be calculated as (1 + 1 + 1 + 2)/4 = 1.25. This formula can also be extended to calculate the MDD of an entire text or treebank, as shown in the adapted formula below. In this adapted formula, n is the total number of words and s is the number of sentences in the text. DDi is the i-th syntactic link of the text or in the treebank.

$$MDD(treebank) = \frac{1}{n-s}\sum_{i=1}^{n-s}|DD_i|$$

The concepts of head-initial and head-final relations are crucial in understanding the notion of dependency direction, which refers to the linear order of two grammatically related words in a sentence (Liu, 2010; Wang and Liu, 2013). Dependency direction has been extensively employed in language typological studies, as it is believed that different

languages exhibit varying tendencies towards head-final or head-initial structures (Chen and Gerdes, 2017; Jiang and Liu, 2015; Liu et al., 2009; Liu, 2010). Moreover, there is a growing trend to apply the concept of dependency direction beyond language typology studies, to explore unique features of diverse language varieties, including those found in translation and interpreting (Fan and Jiang, 2019; Liang and Sang, 2022). Liu (2009) introduced a method for obtaining a frequency distribution of head-initial and head-final dependencies, as demonstrated below. Since a sentence typically only contains two types of dependency relations, head-initial and head-final, the sum of head-final and head-initial frequency distribution in a sentence should be 100% (Liu, 2010; Wang and Liu, 2013). For example, the sentence in Fig. 1 has three head-final grammatical pairs and one head-initial pair. Therefore, the percentage of head-final dependency is 75%, while the head-final percentage is 25%.

$$\text{Percentage of head} - \text{final dependency} = \frac{frequenciesofthehead - finaldependency}{totalnumberofdependenciesinthetreebank} \text{ x } 100$$

$$\text{Percentage of head} - \text{initial dependency} = \frac{frequenciesofthehead - initialdependency}{totalnumberofdependenciesinthetreebank} \text{ x } 100$$

In this study, the transcripts of all three sub-corpora were annotated for their dependency relations using the Stanford Parser, a neural-network powered dependency parser that reveals the grammatical structure of a sentence (Chen and Manning 2014). The output of the automatic parsing was carefully reviewed and manually checked for accuracy before being processed in an Excel format to calculate the MDD and dependency direction. The results are presented in the following section.

4. RESULTS

4.1. Comparing the MDD of the three sub-corpora

The MDD of the texts in the three sub-corpora was calculated and is presented in Fig. 2 below. A comprehensive summary of the statistical information of the results of these calculations is provided in Table 2. The MDDs of the three sub-corpora exhibit similar variation patterns, with values fluctuating between 2.31 and 2.88. Moreover, the mean MDD values of the three sub-corpora are comparable, with small margins of difference: L1O exhibits the highest value (M = 2.61, SD = 0.10), followed by L2O (M = 2.57, SD = 0.12), and L2I (M = 2.55, SD = 0.14). The findings suggest that the syntactic structures of the three language varieties, despite their distinct characteristics, are quite analogous in terms of their level of complexity.

To investigate whether the variation in MDD among the three sub-corpora follows a similar pattern from a statistical perspective, a one-way analysis of variance (ANOVA) test was conducted. Normal distribution of residuals and homogeneity of variances were confirmed, as the number of dependent variables in each group ($N = 50$) exceeded the threshold of 30. The results of the one-way ANOVA revealed that although the MDDs of the three sub-corpora shared the same range of fluctuations, their variation patterns were statistically significant ($F = 3.59$, $p < 0.05$). Tukey's post-hoc tests were conducted to investigate the differences among the MDDs of the three groups: the results are summarized in Table 3 below. The MDD of L2I significantly differs from that of L1O ($p < 0.05$), while no statistical differences were found between L1O and L2O, or between L2O and L2I.

These findings demonstrate that the syntactic structures of interpreted speech are comparatively simpler than those of original non-interpreted speech, providing support for the simplification hypothesis in the field of interpreting. Moreover, the MDD of L2O was observed to be shorter than that of L1O, albeit not statistically significant, but longer than that of L2I. This suggests the intermediate status of L2 speech in the hierarchy of syntactic complexity: in other words, the syntactic structures of L2 speech appear simplified compared to L1 speech, but relatively more complex than those of interpreted speech. Furthermore, the analysis revealed that the difference between L2O and L2I was smaller than that between L2O and L1O. This implies that despite inherent syntactic structural differences among the three language varieties, in terms of syntactic complexity L2 speech exhibits greater similarity to interpreted speech than to L1 speech.

4.2. Comparing dependency direction of the three sub-corpora

The method proposed by Liu (2010) to determine the dependency direction of the three sub-corpora was employed in the present study. This involved calculating the DD of all the dependency pairs in each sub-corpus, and then identifying the frequencies of positive and negative values. These frequencies were subsequently divided by the total count of dependency relations in each sub-corpus to obtain the distribution percentages, which are presented in Table 4 below. It should be noted that in accordance with previous studies, the distribution percentages of the two dependency relations (i.e., head-final and head-initial) always add up to 100% (Liu, 2010; Wang and Liu, 2013). As revealed in Table 4, there

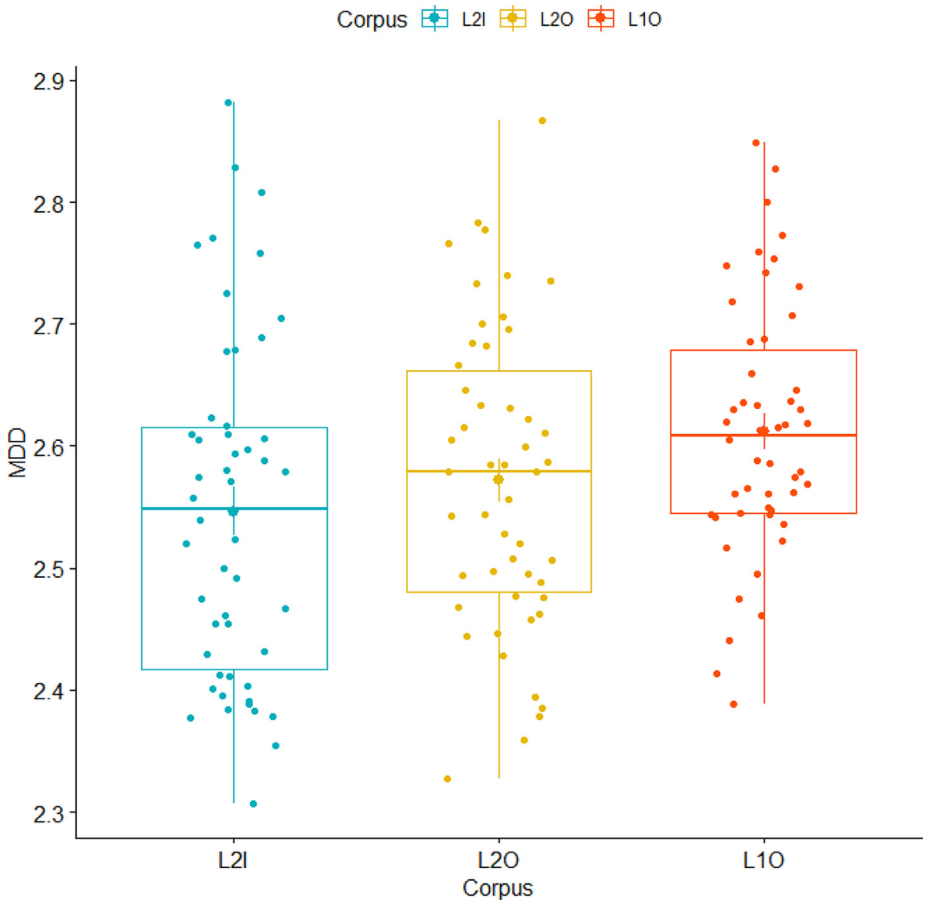

Fig. 2. MDD of the three sub-corpora.

Table 2
An Overview of statistical information of the MDD of the three sub-corpora.

Sub-corpus	Mean	Standard Dev.	Median	Maximum	Minimum
L1O	2.61	0.10	2.61	2.85	2.39
L2O	2.57	0.12	2.58	2.87	2.33
L2I	2.55	0.14	2.55	2.88	2.31

Table 3
Tukey's post-hoc test results of the MDD distinction among the three sub-corpora.

Sub-corpus	Difference	P value
L1O-L2I	0.065	0.025
L2O-L2I	0.025	0.565
L2O-L1O	−0.040	0.242

is a relatively balanced presentation of head-final and head-initial dependencies among the three varieties of English language under investigation. Notably, both interpreted English (54.9%) and L2 English (54.4%) exhibit striking similarity in terms of their dependency direction, with both varieties displaying a higher percentage of head-final dependencies compared to L1 English (51.1%).

Table 4
Dependency direction distribution percentage of the three sub-corpora.

Sub-corpora	Head-final dependency frequency	Head-final distribution percentage	Head-initial dependency frequency	Head-initial distribution percentage	Total dependency frequency
L1O	51,831	51.1%	49,480	48.9%	101,311
L2O	54,708	54.4%	45,854	45.6%	100,562
L2I	54,915	54.9%	45,168	45.1%	100,083

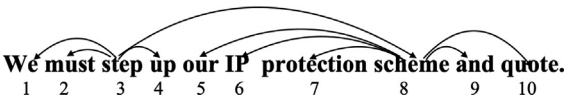

Fig. 3. Dependency relations of sample sentence “We must step up our IP protection scheme and quote” in L2I.

Fig. 4. Dependency relations of sample sentence “There would be no additional cost for business” in L1O.

Previous typological studies have established that English exhibits a relatively balanced distribution of head-final and head-initial relations, while Chinese is characterized as a head-final language (Liu, 2008, 2010). In light of this, the observed increase in head-final dependencies in interpreted English and L2 English may plausibly be attributed to source language interference, whereby the linguistic patterns of the source language influence the syntactic structure of the translational language (Mauranen, 2004; Toury, 2004). An illustrative manifestation of such source language interference can be discerned in the distinctive approach to positioning noun modifiers in Chinese and English (Liang and Sang, 2022). In Chinese, it is usual to place noun modifiers before the noun, resulting in a higher frequency of head-final dependencies in a sentence (Wang and Liu, 2013). This source language practice may be reflected in English speech that is interpreted from Chinese, as exemplified in Fig. 3 below, which presents the dependency structure of a sample sentence obtained from L2I. For instance, in the nominal phrase “our IP protection scheme,” all the modifiers (“our”, “IP”, “protection”) precede the noun (“scheme”), yielding a pure head-final dependency structure. In contrast, the English language tends to exhibit greater flexibility in the positioning of noun modifiers compared to Chinese, as documented in previous studies (Wang and Liu, 2013; Liang and Sang, 2022). English allows for modifiers to be placed both before and after the noun, thereby increasing the potential for head-initial dependency structures. This is exemplified in Fig. 4, where a sample sentence obtained from L1O is analysed for its dependency relations. In the nominal phrase *no additional cost for business*, two modifiers, *no* and *additional*, are positioned before the noun, while one modifier is positioned after the noun connected by the preposition *for*. As a result, this dependency structure gives rise to a combination of two head-final and two head-initial dependency relations, which results in a more balanced distribution compared to the pure head-final dependency structure observed in the English nominal phrase interpreted from Chinese *our IP protection scheme*.

5. DISCUSSION

5.1. Syntactic simplification in interpreted and L2 speech

In this study, a comprehensive corpus consisting of Cantonese-to-English interpreted speech (L2I), original English speech produced by native speakers (L1O), and original English speech generated by proficient non-native speakers (L2O) was employed to examine the dependency relation patterns manifested in these three language varieties. The primary objective was to examine the simplification hypothesis in interpreting, and elucidate the cognitive implications associated with this phenomenon. The empirical findings derived from this study reveal that L2I exhibits the lowest

mean dependency distance (MDD) in comparison to L2O and L1O. This observation underscores the fact that interpreted speech as a whole is characterized by greater syntactic simplification compared to original non-interpreted speech, irrespective of whether it is produced by native or non-native speakers. This evidence is consistent with previous studies which identified simplification in interpreting through lexical measures (Bendazzoli and Sandrelli, 2005; Bernardini et al., 2016; Xu and Li, 2022), and reinforces the growing body of literature examining syntactic simplification in interpreting using measures of syntactic complexity (Liu et al., 2023). In addition, our study shows that, when contrasted with L1O, L2O displays a certain degree of syntactic simplification. This finding suggests that non-native speakers tend to utilize less complex syntactic structures when communicating in English compared to native speakers. As a result, both L2 speech and interpreted speech exhibit similarities in terms of their levels of syntactic complexity. Such an observation aligns with the findings presented by De Clercq and Housen (2017), who demonstrated that L2 French speakers consistently utilized simpler syntactic structures in comparison to native French speakers, including shorter noun phrases and a diminished number of subordinate clauses. The cumulative results of these investigations imply that L2 speakers strategically depend on simpler syntactic structures as a means of compensating for their limited linguistic proficiency in the target language. This compensatory strategy appears to be a pervasive aspect of second language production and interpreting. This finding supports the proposition advanced by Liu et al. (2023) that mediation, manifested in the form of language constraints, may contribute to a reduction in syntactic complexity during the language production process. It also suggests that syntactic simplification may not be unique to translational language; instead, it is more a property of any language variety that arises in a mediated environment under the influence of language contact (Kotze 2020).

5.2. Convergence and divergence of dependency directions

Our findings substantiate previous research (Liu, 2010; Wang and Liu, 2013) confirming English as predominantly characterized by a head-medial distribution, as opposed to head-initial or head-final structures. Head medial distribution, also known as a balanced distribution, refers to the type of syntactic structure where the head of a phrase has roughly an equal number of dependents on either side, resulting in a more balanced and symmetrical sentence structure. Notably, our research demonstrates this structural propensity is consistently maintained across the three language varieties examined. By comparing these language varieties, our analysis also provides compelling evidence that the convergence of dependency direction is more pronounced in interpreted and L2 speech when compared to native English speech. Specifically, our investigation reveals that interpreted speech displays a higher frequency of head-final structures compared to L1 English, thus suggesting the existence of a typological distinction in word order between translated and original languages (Fan and Jiang, 2019; Liang and Sang, 2022). Similarly, L2 speech also contains more head-final structures than L1 native speech. Such a deviation from L1 speech in syntactic structure can be plausibly attributed to the influence of the source language, Chinese, which exhibits a stronger preference for head-final constructions than English (Liang and Sang, 2022; Liu, 2010).

The influence of the Chinese language on L2 speech and interpreting is crucial to consider, as it sheds light on the relationship between source language characteristics and L2 and interpreted output. Chinese, as a head-final language, shows some impact on learning and interpretation processes for both L2 English speakers and interpreters with L1 Chinese backgrounds. This influence is evidenced by a similar inclination toward head-final dependency relations observed in interpreted and L2 language, further reinforcing the notion that these two language varieties, both representing constrained language varieties, share analogous syntactic characteristics (Liu et al., 2023). Overall, our investigation emphasizes the convergent syntactic features of interpreted speech and L2 speech concerning dependency direction, offering valuable insights into the influence of the L1 on the syntactic structure of translated and L2 English. By delving into the similar linguistic processes underlying translation and second language acquisition, our study contributes to a deeper understanding of the typological distinctions observed across different language varieties and the influence of the source language on target language acquisition and interpretation.

5.3. Dependency distance minimisation in interpreted speech and its cognitive implications

The operationalization of dependency relations in a sentence not only serves as a measure of syntactic complexity, but also reflects the working mechanism of human memory. Previous studies have shown that a longer DD leads to greater cognitive demand in the process of language production (Liu, 2008). For example, the DD of translated texts is found to be longer than that of the texts produced in the original language, suggesting that translation activity is more cognitively demanding than native language production (Fan and Jiang 2019; Liang and Sang 2022). The human parser's working memory and ability to cope with cognitive demands are not unlimited, motivating him/her to minimize dependency distance to achieve effective communication without experiencing processing difficulty (Futrell et al., 2015;

Liu et al., 2016; Liu, 2008). This practice of dependency distance minimization (DDM) is supported by empirical studies showing that most sentences or texts in human languages have a threshold DD (Liu, 2008).

Among the three types of language varieties examined in the present study, simultaneous interpreting has long been considered to be a highly cognitively demanding task as it requires interpreters to perform a set of cognitive tasks simultaneously within a very limited timeframe. Gile's well-known Effort Model (2009, p. 168) breaks the process of simultaneous interpreting into a series of sub-tasks, including listening and analysis, short-term memory, and speech production and coordination. When performed simultaneously, all these sub-tasks may give rise to new constraints that do not emerge or have negligible impact when performed individually (Seeber, 2011, p. 186). In addition, the interpreters in the present study needed to work into their L2, which may generate extra cognitive demands (Christoffels et al., 2006). The combined constraints of simultaneous interpreting, coupled with temporal pressure, may motivate interpreters to opt for simpler syntactic structures to reduce cognitive demands, leading to a shorter MDD (Liang et al., 2017). Seen from this perspective, the shortest MDD observed in the case of L2I does not indicate that interpreting activity is less cognitively demanding than original language production. Such a result contrasts with previous research that suggested translated texts have a longer MDD than original texts, implying greater cognitive effort during translation (Fan and Jiang 2019; Liang and Sang 2022). The shorter MDD observed in L2I can be attributed to the operation of the dependency distance minimization principle, which suggests that interpreting is cognitively more demanding than monolingual language production. This finding is consistent with Bóna and Bakti's (2020) observation that interpreting imposes greater cognitive demands than monolingual speech production, as evidenced by the presence of more disfluency markers in interpreting output. It also largely supports the findings of previous experimental studies of the cognitive aspects of interpreting activity (e.g., Chen, 2017; Seeber, 2013) by operationalising the cognitive demand via syntactic dependency features. This result suggests that the corpus-based approach which measures the cognition from a product-oriented perspective can be vital complementary in corroborating the findings of process-focused experimental studies. Moreover, the consumption of cognitive resources in producing different types of language varieties can also be understood in terms of their communication purpose. While both L1 and L2 speakers aim to express their intentions effectively through language, the goal of the simultaneous interpreter is to achieve accurate and complete transfer of another person's message. Given the limits on working memory and temporal constraints in interpreting, interpreters may choose to allocate more cognitive resources to memorizing the content of the message rather than retaining its syntactic complexity. Therefore, syntactic simplification can be seen not only as a strategy to cope with the highly cognitively demanding environment in simultaneous interpreting but also as reflecting the interpreter's decision as to the allocation of cognitive resources in order to meet their professional goal in that particular context.

L2 speech production is often characterized by the presence of an "interlanguage" between the speaker's native and target language (Ouyang and Jiang, 2017; Selinker, 1972). This is reflected in the shorter MDD observed in L2 output in the present study, which suggests that L2 speech production consumes more cognitive resources than native language production. This is understandable, as L2 speakers in the present study need to address the inherent syntactic asymmetry between Chinese and English, two typologically distant languages, and constantly monitor the quality of their utterances to ensure the production of sensible and coherent speech. These constraints are further exacerbated in an oral context where there is no planning time available and L2 speakers are under pressure to produce an utterance immediately (Trebits, 2014). In contrast, for native speakers the process of linguistic encoding in formulating sentences is highly automated, thanks to years of extensive exposure to their first language (Liu et al., 2023; Trebits, 2014). Native speakers do not need to split their attention between the various tasks L2 speakers have to handle, which gives them more available cognitive resources to complexify the syntactic structures and refine the quality of their output (Vercellotti, 2019). In this sense, both L2 speakers and interpreters need to address the syntactic asymmetry between two different languages and produce utterances within a limited timeframe. Consequently, the syntactic structure of L2 speech approximates that of interpreted speech, as evidenced by their MDD and dependency direction. Compared to native speakers, L2 speakers are under greater pressure to simplify their syntactic structure to reduce the cognitive consumption involved in L2 speech production. This finding largely supports previous research by revealing that the level of speech complexity is likely to be reduced when there is an increase in the cognitive load in the process of speech production (Liu et al., 2023; Trebits, 2014; Vercellotti, 2019).

6. CONCLUSION

Situated within the framework of dependency grammar, the present study sets out to investigate the relationship between universal properties of interpreting and human cognitive mechanisms by examining the syntactic structures of interpreted speech in comparison to L1 and L2 speech. Specifically, the present study adopts dependency distance and dependency direction as two measures of dependency relation to examine the level of syntactic complexity, amount of cognitive demand, and word order typological properties, in order to characterize the syntactic structures of the three

language varieties using a monolingual comparative approach (Laviosa, 1998; Baker, 1993). The findings show that interpreters tend to opt for simplified syntactic structures, which feature a lower MDD than native English speakers. The presence of syntactic simplification in interpreting reveals that interpreting is a more cognitively demanding activity than monolingual speech production, which warrants interpreters in reducing dependency distance in order to relieve the cognitive load imposed on their working memory (Liu et al., 2017). In this sense, the phenomenon of simplification can be seen as reflecting the interpreter's efforts to balance the need to produce an accurate rendition in order to fulfil their communication goal and the inevitable influence imposed by cognitive constraints. This result largely corroborates the findings of previous studies that examine simplification at the lexical level (Bendazzoli and Sandrelli, 2005; Bernardini et al., 2016; Xu and Li, 2022). It also confirms the validity of using MDD and dependency direction as reliable syntactic indicators to systematically distinguish translational language from other language varieties by identifying unique syntactic regularities and patterns (Fan and Jiang, 2019; Liang and Sang, 2022). In addition, the present study reveals that L2 speech, as an interlanguage between the source and target speech, is more similar to interpreted speech than to L1 speech in terms of dependency relations characters. This is likely to be a result of the various demanding cognitive tasks involved in the process of L2 speech production, which is akin to interpreted speech production.

The present study extends the investigation of simplification in interpreting from the lexical to the syntactic level by means of the dependency relation measures, enriching existing approaches to exploring the distinctive properties of interpreted speech. However, the study has some limitations. Genre is a factor that affects dependency characteristics, including the length of dependency distance and the fraction of head-initial and head-final relations (Liu, 2008; Wang and Liu, 2017). The present study only examines one genre, and future research may explore how genre can affect the operationalization of dependency features and their demonstration of simplification in interpreting. In addition, simplification is a complex construct, the manifestation of which may be affected by various factors, such as interpreting directionality, language combinations (Bernardini et al., 2016; Lv and Liang 2019), and may take place at different linguistic levels. The present study only examines one interpreting direction, Chinese-English simultaneous interpreting, and measures the manifestation of simplification at the syntactic level. More research efforts should be dedicated to the examination of the potential impact of these influencing factors on simplification and the investigation of their manifestation across various linguistic levels. Moreover, future research can utilize alternative information-theoretic metrics, such as entropy (Liu et al., 2022a, 2022b), to not only validate the existing findings but also enhance our understanding of the distinctive characteristics inherent in interpreted language.

Data availability

Data concerning the study are publicly available on Open Science Framework (https://osf.io/h9dvy

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

The authors would like to extend their sincere gratitude to the Editor and anonymous reviewers for their insightful comments and suggestions.

References

Anderman, G., Rogers, M., 2008. The linguist and the translator. Incorporating corpora: Linguist Translator 2, 5.

Baker, M., 1993. Corpus linguistics and Translation Studies: Implications and applications. In: Baker, M. (Ed.), Text and Technology: In Honour of John Sinclair. Benjamins, Amsterdam, pp. 233–250.

Bendazzoli, C., 2018. Corpus-based interpreting studies: Past, present and future developments of a (wired) cottage industry. Making Way Corpus-based Interpreting Stud., 1–19.

Bendazzoli, C., Sandrelli, A., 2005. An approach to corpus-based interpreting studies: developing EPIC (European Parliament Interpreting Corpus). Proc. Challenges Multidimensional Transl..

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.

Bialystok, E., Craik, F.I., 2010. Cognitive and linguistic processing in the bilingual mind. Curr. Dir. Psychol. Sci. 19 (1), 19–23.
Bóna, J., Bakti, M., 2020. The effect of cognitive load on temporal and disfluency patterns of speech: evidence from consecutive interpreting and sight translation. Target 32 (3), 482–506.
Chen, S., 2017. The construct of cognitive load in interpreting and its measurement. Perspectives 25 (4), 640–657.
Chen, X., Gerdes, K., 2017. Classifying languages by dependency structure. Typologies of delexicalized universal dependency treebanks. In: Proceedings of the fourth international conference on dependency linguistics (Depling 2017), pp. 54–63.
Chen, D., Manning, C.D., 2014. A fast and accurate dependency parser using neural networks. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 740–750.
Chesterman, A., 2010. Why study translation universals. Acta Translatol. Helsingiensia 1, 38–48.
Christoffels, I., De Groot, A., 2009. Simultaneous interpreting. Handbook of bilingualism: Psycholinguistic approaches, 454–479.
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.
Christoffels, I.K., De Groot, A.M., Kroll, J.F., 2006. Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. J. Mem. Lang. 54 (3), 324–345.
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.
De Sutter, G., Lefer, M.A., 2020. On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach. Perspectives 28 (1), 1–23.
Fan, L., Jiang, Y., 2019. Can dependency distance and direction be used to differentiate translational language from native language? Lingua 224, 51–59.
Ferraresi, A., Bernardini, S., Petrović, M.M., Lefer, M.A., 2018. Simplified or not simplified? The different guises of mediated English at the European Parliament. Meta 63 (3), 717–738.
Futrell, R., Mahowald, K., Gibson, E., 2015. Large-scale evidence of dependency length minimization in 37 languages. Proc. Natl. Acad. Sci. 112 (33), 10336–10341.
Gibson, E., 1998. Linguistic complexity: Locality of syntactic dependencies. Cognition 68 (1), 1–76.
Gile, D., 2009. Basic Concepts and Models for Interpreter and Translator Training. John Benjamins Publishing.
Grodner, D., Gibson, E., 2005. Consequences of the serial nature of linguistic input for sentenial complexity. Cognit. Sci. 29 (2), 261–290.
Gumul, E., 2021. Explicitation and cognitive load in simultaneous interpreting: product-and process-oriented analysis of trainee interpreters' outputs. Interpreting 23 (1), 45–75.
Hudson, R., 2010. An Introduction to Word Grammar. Cambridge University Press.
Jiang, X., Jiang, Y., 2020. Effect of dependency distance of source text on disfluencies in interpreting. Lingua 243, 102873.
Jiang, J., Liu, H., 2015. The effects of sentence length on dependency distance, dependency direction and the implications–based on a parallel English-Chinese dependency treebank. Lang. Sci. 50, 93–104.
Kajzer-Wietrzny, M., 2012. Interpreting Universals and Interpreting Style. Adam Mickiewicz University, Poznań, Poland, Unpublished PhD dissertation..
Kajzer-Wietrzny, M., 2013. Idiosyncratic features of interpreting style. New Voices in Translation Studies 9 (1), 38–52.
Kotze, H., 2020. Translation, contact linguistics and cognition. In: The Routledge Handbook of Translation and Cognition. Routledge, pp. 113–132.
Kruger, H., Van Rooy, B., 2016. Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. English World-Wide 37 (1), 26–57.
Laviosa, S., 1998. Core patterns of lexical use in a comparable corpus of English narrative prose. Meta 43 (4), 557–570.
Lei, L., Wen, J., 2020. Is dependency distance experiencing a process of minimization? A diachronic study based on the State of the Union addresses. Lingua 239, 102762.
Levon, E., Buchstaller, I., 2015. Perception, cognition, and linguistic structure: the effect of linguistic modularity and cognitive style on sociolinguistic processing. Lang. Var. Chang. 27 (3), 319–348.
Liang, J., Fang, Y., Lv, Q., Liu, H., 2017. Dependency distance differences across interpreting types: implications for cognitive demand. Front. Psychol. 8, 2132.
Liang, Y., Sang, Z., 2022. Syntactic and typological properties of translational language: a comparative description of dependency treebank of academic abstracts. Lingua 273, 103345.
Liu, H., 2008. Dependency distance as a metric of language comprehension difficulty. J. Cognit. Sci. 9 (2), 159–191.
Liu, H., 2009. Probability distribution of dependencies based on a Chinese dependency treebank. J. Quant. Linguist. 16 (3), 256–273.
Liu, H., 2010. Dependency direction as a means of word-order typology: a method based on dependency treebanks. Lingua 120 (6), 1567–1578.
Liu, K., Afzaal, M., 2021. Syntactic complexity in translated and non-translated texts: a corpus-based study of simplification. PLoS One 16 (6).
Liu, Y., Cheung, A.K., Liu, K., 2023. Syntactic complexity of interpreted, L2 and L1 speech: a constrained language perspective. Lingua 286, 103509.
Liu, H., Hudson, R., Feng, Z., 2009. Using a Chinese treebank to measure dependency distance. Corpus Linguist. Linguist. Theory 5 (2), 161–174.

Liu, K., Liu, Z., Lei, L., 2022a. Simplification in translated Chinese: an entropy-based approach. Lingua 275, 103364.
Liu, K., Ye, R., Liu, Z., Ye, R., 2022b. Entropy-based discrimination between translated Chinese and original Chinese using data mining techniques. PLoS One 17 (3), e0265633.
Liu, M., Schallert, D.L., Carroll, P.J., 2004. Working memory and expertise in simultaneous interpreting. Interpreting. Int. J. Res. Pract. Interpret. 6 (1), 19–42.
Liu, H., Xu, C., Liang, J., 2016. Dependency length minimization: puzzles and promises. Glottometrics 33, 35–38.
Liu, H., Xu, C., Liang, J., 2017. Dependency distance: a new perspective on syntactic patterns in natural languages. Phys. Life Rev. 21, 171–193.
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.
Ma, X., Cheung, A.K., 2020. Language interference in English-Chinese simultaneous interpreting with and without text. Babel 66 (3), 434–456.
Mauranen, A., 2004. Corpora, universals and interference. *Translation Universals–Do they exist*, 65–82.
Neumann, S., Serbina, T., 2020. Translation, corpus linguistics and cognition. In: The Routledge Handbook of Translation and Cognition. Routledge, pp. 188–205.
Ouyang, J., Jiang, J., 2017. Can the probability distribution of dependency distance measure language proficiency of second language learners? J. Quant. Linguist. 25 (4), 295–313.
Pym, A., 2008. On Toury's laws of how translators translate. Benjamins Translation Library 75, 311.
Rennert, S., 2008. Visual input in simultaneous interpreting. Meta 53 (1), 204–217.
Rinne, J.O., Tommola, J., Laine, M., Krause, B.J., Schmidt, D., Kaasinen, V., Sunnari, M., 2000. The translating brain: cerebral activation patterns during simultaneous interpreting. Neurosci. Lett. 294 (2), 85–88.
Russo, M., Bendazzoli, C., Sandrelli, A., 2006. Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: extended analysis of EPIC (European Parliament Interpreting Corpus). Forum 4 (1), 221–254.
Seeber, K.G., 2011. Cognitive load in simultaneous interpreting: Existing theories—new models. Interpreting 13 (2), 176–204.
Seeber, K.G., 2013. Cognitive load in simultaneous interpreting: measures and methods. Target. Int. J. Transl. Stud. 25 (1), 18–32.
Seeber, K.G., Kerzel, D., 2012. Cognitive load in simultaneous interpreting: model meets data. Int. J. Biling. 16 (2), 228–242.
Selinker, L., 1972. Interlanguage. *Int. Rev.* Appl. Linguis. 10, 209–231.
Shen, M., Lin, Y., Lv, Q., Liang, J., 2023. A corpus-based analysis of the effect of syntactic complexity on disfluency in consecutive interpreting. Lingua 391, 1–14.
Shlesinger, M., 1998. Corpus-based interpreting studies as an offshoot of corpus-based translation studies. Meta 43 (4), 486–493.
Shlesinger, M., 2000. Interpreting as a Cognitive Process: How Can We Know What Really Happens? Benjamins Translation Library, pp. 3–16.
Shlesinger, M., Ordan, N., 2012. More spoken or more translated? Exploring a known unknown of simultaneous interpreting. Target. Int. J. Transl. Stud. 24 (1), 43–60.
Stefanowitsch, A., Flach, S., 2017. The corpus-based perspective on entrenchment.
Swabey, L., Nicodemus, B., Taylor, M.M., Gile, D., 2016. Lexical decisions and related cognitive issues in spoken and signed language interpreting: a case study of Obama's inaugural address. Interpreting. Int. J. Res. Pract. Interpret. 18 (1), 34–56.
Teich, E., 2003. Cross-linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts. de Gruyter.
Toury, G., 2004. Probabilistic explanations in translation studies. Translation universals: Do they exist.
Trebits, A., 2014. Sources of individual differences in L2 narrative production: the contribution of input, processing, and output anxiety. Appl. Linguis. 37 (2), 155–174.
Vercellotti, M.L., 2019. Finding variation: assessing the development of syntactic complexity in ESL Speech. Int. J. Appl. Linguist. 29 (2), 233–247.
Wang, L., Liu, H., 2013. Syntactic variations in Chinese-English code-switching. Lingua 123, 58–73.
Wang, Y., Liu, H., 2017. The effects of genre on dependency distance and dependency direction. Lang. Sci. 59, 135–147.
Xu, C., Li, D., 2022. Exploring genre variation and simplification in interpreted language from comparable and intermodal perspectives. Babel 68 (5), 742–770.
Yan, J.X., Liang, J., 2022. Foreign language anxiety and dependency distance in English-Chinese interpretation classrooms. Front. Psychol. 13, 1–10.