This is the Pre-Published Version.

This is the accepted version of the publication Xu, C., & Li, D. (2022). Exploring genre variation and simplification in interpreted language from comparable and intermodal perspectives. Babel, 68(5), 742-770. The Version of Record is available online at: https://doi.org/10.1075/babel.00289.cui.

Exploring genre variation and simplification in interpreted language from comparable and intermodal perspectives

Xu Cui^a and Li Dechao^b

a Beijing Institute of Technology b The Hong Kong Polytechnic University

This article explores genre variation and simplification in interpreted language from both comparable (interpreted vs. non-interpreted/nonmediated) and intermodal (interpreted vs. translated) perspectives. It draws on a newly built unidirectional comparable and intermodal corpus named the LegCo+, which features legislative proceedings in the Legislative Council of Hong Kong (including originals and their translations and interpretations from Cantonese to English), as well as original plenary speeches delivered by native English speakers in the UK Parliament. It investigates the variation patterns of four simplification features in three dimensions, namely, standardized type-token ratio (STTR) and list heads for lexical diversity, lexical density for informativeness, and average sentence length for lexical sophistication. It aims to uncover the effects of mediation and genre, as well as their interaction effects on linguistic variation. The results indicate that texts of different mediation statuses and genre categories vary with respect to simplification patterns. From a comparable perspective, interpretations rely on a narrower range of vocabulary than non-interpretations, but they are also more informative, and such informativeness is dependent on genre categories. Intermodally speaking, interpretations exhibit consistent patterns of simplification, indicating a strong modality (or mode of mediation) effect.

Keywords: genre variation, simplification, comparable, intermodal, variation patterns

1. Introduction

Over the past few decades, there has been abundant empirical evidence demonstrating the overall more simplified nature of translated language in relation to target originals. Generally speaking, translated language is reported to feature less varied and more repetitive lexicons, higher frequency words, shorter mean sentence length, or simply, to make "do with less words" (Blum-Kulka and Levenston 1978, 399). Previously, this pattern of "lexical simplification" was believed to be translation-inherent (Baker 1993). Like many other features of translation (such as explicitation and normalization), they are "almost the inevitable by-products of the process of mediating between two languages rather than being the result of the interference of one language with another" (Laviosa 2002, 43). Recently, however, much effort has been made to disentangle the possible contributors to this mystery, of which genre, or register in a broader sense, stands out as a main factor (Kruger and Van Rooy 2012; Hu et al. 2016). Empirical investigations have revealed that while certain features show ontological differences between translations and non-translations in one genre, they do not necessarily exhibit a statistically significant difference in a different genre. Laviosa (1998) reports, for example, that mean sentence length is shorter in translated newspaper articles, but an opposite trend characterizes translated narrative prose, which highlights possible genre variation.

Compared with the rigorous research on features of translation, interpreting has only recently begun to receive academic attention in terms of its distinctive features as compared to non-mediated (i.e., native) spoken language and, more recently, translated language (see Section 2.1). However, such interlingual comparable comparisons report mixed findings regarding lexical simplification in interpreting, with language pair/combination, working direction, and source language influence being the most influencing factors (Sandrelli and Bendazzoli, 2005; Russol et al., 2006). To our best knowledge, no research on genre/register variation has been carried out due to the inherent difficulties in collecting data from various settings and the predominance of interpreting data being based on debate proceedings from the European Parliamentary setting (Sandrelli and Bendazzoli, 2005; Bernardini et al. 2016).

In this study, we will focus on genre variation and lexical simplification in interpreted language from both comparable and intermodal perspectives. Following Biber (1988, 170), we define genre based on "the external criteria relating to the speaker's purpose and topic; [the genre categories] are assigned on the basis of use rather than on the basis of form." Applied to our study, we have categorized our corpus data into two general genre categories, namely, *Questions* & *Answers* sessions (i.e., Q&As) and *Debates*. The external criteria relating to the speaker's purpose and topic are different for the two genre categories: while Q&As aim to offer information concerning questions raised, *Debates* aim to persuade parliamentary members and negotiate motions for change. However, to make a more fine-grained analysis, we've further categorized the Q&As into two subgenres, coded as *Q&As_type1* and *Q&As_type2* (see Section 3 for a more detailed explanation).

The inclusion of two different perspectives is based on the legacy of Shlesinger (2008) and Shlesinger and Ordan (2012), which view interpreting as both *spoken* and *translated* (we prefer the term "mediated") discourse. We are interested in discovering whether simplification features can be identified in interpreted texts as compared to non-mediated and translated texts from both comparable and intermodal perspectives. This can be operationalized as genresensitivity in features such as standardized type-token ratio (STTR), list heads, lexical density, average sentence length, and high-frequency words. That is, we aim to explore the following two broad research questions: (1) what are the variation patterns of simplification features in texts of three mediation statuses (i.e., non-mediated, interpreted, and translated)? (2) what is the relationship between genre and simplification features in interpreting?

Based on the two research questions, we aim to investigate the effect of mediation status (i.e., non-mediated, interpreted, and translated texts) and genre categories (i.e., $Q & As_type1$, $Q & As_type2$, and Debates) on the variation patterns of simplification features, based on a newly built intermodal comparable corpus, which we have called the LegCo+. We are also interested in observing the interaction effects between mediation and genre to uncover the potential relationship between the two. Our initial hypotheses are that: (1) texts of different mediation status will demonstrate statistically significant differences in terms of the variation patterns of simplification features, with interpreting standing out from the other two given its special status of being both spoken and mediated discourse; (2) texts of different genre categories will demonstrate statistically significant differences in terms of the variation patterns of simplification; and (3) there will be no significant interaction effects between mediation and genre, in that mediated texts (including translations and interpretations) will show homogeneous patterns and exhibit a "leveling out" effect (Baker 1993).

2. Literature review

2.1 Simplification in translation and interpreting

The first academic proposal on the simplification hypothesis was made by Blum-Kulka and Levenston (1978, 399), who define simplification as "the process and/ or result of making do with less words" in language learning. Later, Baker (1993) includes this hypothesis as one of the four well-known translation universals, i.e., simplification, explicitation, normalization, and leveling out, and proposes a

new research paradigm that compares translated texts with non-translated original texts based on comparable corpora. Following the Bakerian paradigm, many translation scholars set out to verify whether translations are more simplified than non-translations in terms of lexical variety, lexical density, list head coverage, high-frequency words, hapax legomena, mean sentence length, and others (Laviosa 1997, 1998, 2002; Williams 2005; Xiao 2010; Xiao and Hu 2015). Amongst these studies, Laviosa's (1997, 1998) groundbreaking works on simplification patterns in translation have been most replicated by interpreting scholars. Based on a comparable analysis, she (1998, 8) finds four core patterns of lexical simplification in translational versus non-translational texts, which include:

- (1) Translated texts have a relatively lower percentage of content words versus grammatical words (i.e., their lexical density is lower);
- (2) The proportion of high-frequency words versus low-frequency words is relatively higher in translated texts;
- (3) The list head of a corpus of translated texts accounts for a large area of the corpus (i.e., the most frequent words are repeated more often);
- (4) The list head of translated texts contains fewer lemmas.

These lexical patterns exist not only in translations of narrative prose but also in those of newspaper articles and thus can be regarded as the core patterns of lexical use in translation. However, there is one linguistic feature that is not shared in translated newspaper articles and narrative prose, i.e., mean sentence length. Laviosa (1998) reports shorter mean sentence length in translated news while longer sentence length in translated literary works, indicating possible genre variation in this regard. Besides this contradictory finding, there are other studies offering counterevidence to the simplification hypothesis, such as more frequent use of untypical collocations (Mauranen 2000), more frequent use of modifiers (Jantunen 2001), longer mean sentence length (Xiao and Dai 2014), and higher type-token ratio (Qin and Wang 2009).

Research on simplification in interpreting, by contrast, dates back to Sandrelli and Bendazzoli (2005) on the lexical patterns of simultaneously interpreted languages. Following Laviosa (1998); Sandrelli and Bendazzoli (2005) investigate lexical density and list heads in four language combinations, i.e., Spanish/Italian to English, and Spanish/English to Italian, based on the European Parliament Interpreting Corpus (EPIC). Their results lend limited support to simplification in interpreting, and language pair/combination turns out to be an influencing factor. For lexical density, there is little variation in original and interpreted English, while lexical density in interpreted Italian from Spanish and from English shows opposite trends. For list heads, once again, they report contradictory findings in interpreted English and Italian, with the former confirming lexical simplification while the latter not. The mixed findings reported indicate that conclusions drawn from written translations cannot be readily applied to interpreting and language pairs, and working direction may have a direct influence on linguistic variation in interpreting.

Following Sandrelli and Bendazzoli (2005); Russo et al. (2006) carry out an extended analysis of the EPIC corpus on the lexical patterns of interpreted Spanish from English and Italian respectively, in relation to original Spanish. The same indicators for lexical simplification are investigated, i.e., lexical density and list heads. They also report contradictory findings: while interpreted Spanish is characterized by higher lexical density compared with original Spanish, it also features high-frequency word usage. Similar contradictory results have also been reported in interpreting from Chinese into English: while Hu and Tao (2010) and Li and Wang (2012) identify simplification patterns of interpreted English in list heads and lexical density, the opposite trends have been reported in Chen and Cui (2010). In a similar vein, Kajzer-Wietrzny's (2012, 2015) comparable analysis also reports mixed patterns as regards interpreting, as interpreted English is found to be more lexically dense and informative, and at the same time more repetitive than non-interpretations. In comparison, translational English conforms to the simplification hypothesis in all three parameters (i.e., lexical density, core vocabulary, and list heads).

A more recent study by Dayter (2018) looks into lexical variety (i.e., STTR) and lexical density in a newly built interpreting corpus called SIREN, a parallel aligned bidirectional corpus of original and simultaneously interpreted speech in Russian and English. Her analysis also reports mixed results: while the Russian sub-corpora have generally confirmed the simplification trend, the opposite is true for the English counterpart. She argues that one possible reason might be working direction since the Russian-to-English subcorpus "consists of up to a third of samples from interpreters working into their B language" (Dayter 2018, 257), which may lead to over-correction of simultaneous interpreters, leading eventually to higher lexical density and more elaborate and varied vocabulary. However, we argue the opposite as we believe that if the working direction does play a role, it is more likely that interpreted language from A to B should be even more simplified than native interpreting (B-to-A), given the much higher cognitive constraints involved.

2.2 Register/genre variation in translation and interpreting

Early studies on the typical features of translation, as reviewed above, tend to focus on only one register or genre category. Laviosa's (1997) first attempt at lexical patterns of translation focuses on newspaper articles, followed then by

her seminal work (1998) on simplification patterns of translated narrative prose. Her studies show that translations of both genres are characterized by a narrower vocabulary range, lower lexical density, and more repetitive language use, and thus she identifies these features as the core lexical patterns of translated language. She (1998, 9) further suggests that these core patterns of lexical use can be tested "on a variety of translational text genres and different types of translation (for example, conference, court interpreting, etc.)" so as to "establish whether and to what extent these regularities are subject field and/or modality-specific and/or language-specific, or whether they can indeed be considered universal features of translational English."

Despite this proposition, very few studies include genre/register variation when investigating typical features of translation. One of the most representative works has been carried out by Kruger and Van Rooy (2012), which examines several linguistic parameters associated with simplification, explicitation, and normalization across six registers in two corpora (translational and nontranslational). Their study offers limited support for the effect of mediation, except for that omission and lexical diversity, whilst it offers strong evidence of register variation as well as interaction effects between register and mediation status. In their concluding remarks, they appeal for more nuanced investigations with respect to the features of translated language since "[t]he distribution and prevalence of linguistic realisations that may be linked to these features of translated language are variable and subject to the influence of a variety of factors, amongst others registers" (Kruger and Van Rooy 2012, 62). A similar concern has also been expressed by Ferraresi et al. (2018, 134) when they claim that "[r]egister and source language variables should therefore be carefully considered when designing studies that aim to detect typical features of translated language."

More recently, Hu et al. (2016), utilizing a more sophisticated multidimensional approach, investigate the features of translation based on two large-scale corpora, i.e., the balanced Corpus of Translational English (COTE) and the Freiburg-LOB corpus of British English (FLOB), covering a wide variety of registers. Although register variation is not their research focus, their analysis does reveal strong register variation since they have identified only 6 out of 96 linguistic features that are "distinctive at the overall level, in both the broad categories and in all four major registers" of which lexical density and top 10 vocabulary are related to simplified language use. In other words, translational English is found to be consistently more simplified than non-translational English in terms of lexical density and top 10 vocabulary across all registers, while the remaining features all demonstrate genre/register variation.

While research on translated language has only recently begun to include genre or register variation, research on interpreting seems to be lagging behind. One possible reason is the much more daunting task involved in constructing an interpreting corpus covering multiple genre categories. One reality of recorded interpretations is that most of the representative interpreting corpora, such as EPIC and Europarl, are based on plenary speeches (mostly debates) in the European Parliament. We argue that, even within the parliamentary setting, there are still proceedings that can be categorized into different genres, such as *Debates* versus QeAs, and differences may exist in terms of the linguistic realizations in their interpreted versions. Failing to examine the specific linguistic manifestations of each speech type, we may have masked their nuances and neglected the possible variations within these texts, which may lead to contradictory findings that are hard to interpret.

The current study attempts to address this concern by investigating simplification features in simultaneously interpreted English (translated from Cantonese) in three genres, with reference to original non-mediated English and translated English from the same source. The motivation is to examine as closely as possible any linguistic variation evident in each speech type (i.e., genre) with respect to simplification.

2.3 Translation, interpreting, and intermodal comparison

While the bulk of research adopts a comparable approach, that is, comparing translations to non-translations to isolate typical features of translation, Shlesinger (1989, 2008) recognizes the special status of interpreting as being both spoken language and a form of translation and proposes a comparable intermodal approach. This method compares interpreted language with both non-interpreted language and translated language from the same or similar source. Her aim was to see whether interpreting exhibits more features of non-mediated spoken language as intrinsically as a segment of oral discourse or whether it exhibits more features of translation as a form of translated discourse (i.e., *mediated* discourse).

Shlesinger's (2008) proposal has been acknowledged by many interpreting scholars. Following her call, Shlesinger and Ordan (2012) examine lexical density, type-token ratio (TTR), and a few part-of-speeches based on small-scale comparable intermodal corpora. Their findings indicate that interpreting is more spoken than translated, and in many cases, interpreting can be regarded as "an extreme case of translation" (ibid., 54), where the typical patterns of translation (in relation to non-translations), such as simplification, are found to be more prominent in interpreting.

Similarly, Bernardini et al. (2016) investigate lexical simplification patterns based on the European Parliament Translation and Interpreting Corpus (EPTIC). Their results have overall confirmed the simplification hypothesis in both interpreting and translation, and they suggest that "the mediation process reduces complexity in both modes of language production and both language directions [English↔Italian], with interpreters simplifying the input more than translators" (ibid., 61). In conclusion, they concur with Shlesinger and Ordan's (2012) view of interpreting as "an extreme case of translation" as far as simplification is concerned.

Other intermodal studies on simplification have been carried out by Ferraresi et al. (2018) on translation versus interpreting, as well as Lü and Liang (2018) comparing consecutive interpreting versus simultaneous interpreting, among others. Ferraresi et al. (2018) do not unconditionally support the simplification hypothesis, which they had operationalized by using the same set of parameters as in Laviosa (1997, 1998). Their study indicates that both the mediation mode (i.e., translation and interpreting) and the source languages have an influence on simplification patterns, with the influence of the latter being greater than the former. Interpreted English from French exhibits overall more simplified patterns, but this is missing in interpreted English from Italian. Nevertheless, lexical simplification is greater in interpretations than translations, although the evidence is also not equally stronger for the two source languages. Moreover, their study does not provide support for lexical simplification in translations, as argued by Laviosa (1997, 1998).

Amongst the mixed findings, one trend seems to be much clearer, that is, interpreted language is more repetitive than non-interpreted language, being characterized by higher list heads and lower type-token ratio. Findings with respect to lexical density and core vocabulary coverage usually offer counterevidence, leading to mixed results, which are often interpreted from different perspectives, such as language combination or working direction, mode of delivery, modes of mediation, and interpreters' strategies.

Despite the mixture of evidence and counterevidence, perhaps the most important thing of these studies is "to consider and compare various factors, supportive or subversive, to reach a more detailed and hence more profound understanding of simplification" (Xiao and Hu 2015, 159), rather than to testify whether there are universals in texts of mediation in an absolute sense. In this study, we attempt to contribute to this aspect by considering the effects of mediation and genre based on a newly built comparable intermodal corpus of translations, interpretations, and non-interpretations in the Cantonese-to-English language direction.

3. Methodology

3.1 Introducing the LegCo+ corpus

Our corpus data are drawn from a newly built million-size comparable intermodal corpus named the LegCo+ corpus, which was specifically designed for the first author's Ph.D. project (see Table 1). It contains four sub-corpora or components: Cantonese source speeches, simultaneously interpreted speeches into English, translated speeches into English, and original English speeches. The corpus includes three genres collected from the proceedings of the Legislative Council of Hong Kong and the House of Commons in the UK Parliament during the year period 2015 to 2017. Out of the four sub-corpora, three are English texts of different mediation status, that is, non-mediated (NS), interpreted (SI), and translated English (WT), and they are the focus of the current study. The three genres include Question and Answer Sessions to the Prime Minister/Chief Executive (labeled as Q&As_type1), Questions to the Ministers/Secretaries (labeled as Q&As_type2), and Debates on motions and bills (labeled as Debates). We are fully aware that, technically, Q&As_type1 and Q&As_type2 belong to the same genre category (Biber 1988), that is, Q&As. However, we decided to categorize the two separately, considering the different production conditions of their source speeches, especially in terms of modes of delivery (i.e., scripted, unscripted, or mixed). For Q&As_type2, they contain both written (written to be read) and spoken Q&As, which is not the same case for Q&As_type1 (which is spoken only). Categorizing them under the same genre category will mask the nuances between the two, leading to possible contradictory findings that are hard to interpret.

Given the uneven time length of each proceeding selected, transcriptions of the three genres were segmented into 477 text segments altogether, with the total running words ranging from 1,300 to 2,000 per text segment (see Table 2).

	ST	SI	WT	NS
Corpus size	400,000 (characters)	235,156 (tokens)	301,292 (tokens)	228,174 (tokens)
Language	Cantonese	English	English	English
Mediation status	Non-mediated (native)	Mediated (non- native)	Mediated (non- native)	Non-mediated (native)
Total time length	28h5m	28h5m	N/A	22h49m

Table 1. Outline of the LegCo+ corpus

Table 1. (continued)

	ST	SI	WT	NS			
Genres	Q&As_type1: Questions and Answers to the Prime Minster/Chief Executive Q&As_type2: Questions to the Ministers/Secretaries; and, Debates: Debates on Motions and Bills						
Setting	The Hong Kong I	egislative C	ouncil	The UK Parliament			
Participants power relations	Chief Executive, 1 Members (includ representing diffe	President, Le ing Secretari rent politica	egislative es) l parties	Prime Minister, Speaker, Parliamentary Members (including Ministers) representing different political parties			
Time period	2015-2017						

English components	Mediation status	Q&As_type1	Q&As_type2	Debates	Total
NS	Non-mediated	49	29	60	138
SI	Interpreted	51	27	71	149
WT	Translated	65	30	95	190
Total		165	86	226	477

Table 2. Text segments included in three genre categories

3.2 Features to be investigated

In the current study, we focus on simplification features from three dimensions: lexical diversity, operationalized by standardized type-token ratio and list heads; informativeness, indexed by lexical density; and syntactic complexity, understood as average sentence length, following previous research traditions (e.g., Bernardini et al. 2016; Kajzer-Wietrzny 2015; Laviosa 1998; Sandrelli and Bendazzoli 2005; Lü and Liang 2018). The reason for such a selection is straightforward, that is, to make a direct comparison with previous findings based on the much less investigated interpreting direction of Cantonese-to-English.

Standardized type-token ratio, or STTR, is a much-investigated parameter for lexical variety or diversity (e.g., Shlesinger and Ordan 2012; Dayter 2018) since it can be used to calculate the percentage of different types of words in a total amount of running words. STTR in this research was automatically calculated using the WordSmith tool v.6.0 (Scott 2012) on the basis of 1,000 words. Given the spoken nature of SI and NS English components, the originally transcribed paralinguistic features, such as truncated words, filled pauses ("erm," "ehm," "uh"), repairs, etc., were removed before the calculation to improve calculation accuracy. Another much-discussed feature for lexical variety/diversity is list heads, which are calculated as the percentage of the top hundred words in the wordlists for each text segment. The wordlists (for each text segment) were also generated by the WordSmith tool v.6 (Scott 2012). As explained by Sandrelli et al. (2010, 185), "if the 100 most frequent words in a sub-corpora account for a large part of that sub-corpus, lexical variety is low because it means that the same words are used over and over again." In other words, higher list heads mean that the texts are highly repetitive and less lexically varied.

In terms of informativeness, we focus on lexical density, which calculates the ratio/percentage between lexical words and the total running words in a corpus (Laviosa 1998; Sandrelli and Bendazzoli 2005; Kajzer-Wietrzny 2015). Since our corpus is PoS-tagged, the number of lexical words (content words) can be easily calculated by summing up the number of nouns, verbs, adjectives, and adverbs annotated by TreeTagger (Schmid 1994), with its reported tagging accuracy reaching 97% to 98%.

The last feature measuring syntactic complexity, or lexical sophistication, is average sentence length. Previous studies (Ferraresi et al. 2018) have argued that average sentence length is not a reliable parameter since it depends on text type and is therefore discarded in many studies. We decided, however, to keep this feature as a possible indicator of lexical sophistication, following Bernardini et al. (2016). For both spoken and written components of the LegCo+ corpus, average sentence length is calculated by dividing the total running words in each text segment by the number of sentences in that text segment. For interpreting and non-mediated speech, punctuation markers were added during the transcription, based on the intonation of the speakers or interpreters as well as syntactic information, which makes the calculation of average sentence length much easier.

We acknowledge the fact that the selected four features are far from reflecting the true and complete picture of the possible simplification trends of translation and interpreting, as there must be many other features associated with simplification, as reported in recent studies utilizing multivariate approaches (Hu et al. 2016; Kruger and Van Rooy 2018) that are currently beyond the scope of discussion for this study. We do believe, however, that a replication study based on genetically distinct language pairs (such as Cantonese and English) will help better reveal the nature of mediated language before we move on to uncover its multifaceted nature.

3.3 Data processing and statistical analysis

As introduced above, the four simplification features operationalized as STTR, list heads, lexical density, and average sentence length were calculated either automatically or manually adjusted for each text segment. Based on correlation analysis, we find a very strong negative correlation between STTR and list heads (r=-.848), suggesting that our operationalization of the two features as indicators for lexical diversity/variety is appropriate. While previous studies (Lü and Liang 2018; Xiao and Yue 2009) claim that lexical density is affected by average sentence length, we find a relatively low correlation coefficient between the two (r=-.0423), which also helps justify our decision to separate the two in two dimensions (lexical density for informativeness and sentence length for syntactic complexity).

For the exploration of main effects and interaction effects, we have applied two-way ANOVA, with mediation (non-mediated, interpreted, and translated) and genre ($Q & As_type1$, $Q & As_type2$, and Debates) as the two independent variables, while frequencies/percentage of the four features have been deployed as the dependent variables. Pairwise comparisons were performed when significant main effects were identified so as to determine if the differences among texts of three mediation statuses are specific to a certain genre but not to others. In all cases, two levels of statistical significance were used as benchmarks, i.e., p < .05 and p < .001, following Kruger and Van Rooy (2012).

4. Results and analysis

4.1 Overview

Table 3 presents an overview of the simplification patterns in the non-mediated, interpreted, and translated speeches. Overall, interpretations show consistent patterns of being more simplified than non-mediated speech delivered by native English speakers, except for lexical density, characterized by a narrower range of vocabulary, lower lexical variety, and much shorter average sentence length. Lexical density shows a contrasting trend, although the difference is not as distinctive as the other indicators. Intermodally speaking, interpretations confirm the simplification hypothesis for the four indicators along all three dimensions, indicating a strong effect of modality or modes of mediation.

Dimensions	Features	NS	SI	WT
Lexical diversity	STTR	39.50%	36.56%	39.29%
	List heads	62.74%	66.21%	63.07%
Informativeness	Lexical density	54.33%	54.97%	56.94%
Syntactic complexity	Average sentence length	20.08	16.80	23.25

 Table 3. Overall variation patterns of simplification features across three mediation status

Table 4 zooms in on the variation patterns of simplification features across all sub-genres.¹ This breakdown analysis shows that genre variation exists even within texts of the same mediation status. For original speech delivered by native English speakers, *Debates* always stand out from the other two types of *Q&As*, featured by lower STTR, higher list heads, and lower lexical density, while surprisingly longer average sentence length. For interpreted speech, the distinctive subgenre is $Q&As_type2$ with respect to STTR and average sentence length. Interpreted *Debates* are also characterized by the lowest lexical density. For translated speech, once again, $Q&As_type2$ stands out for all simplification features, indexing lower STTR, higher list heads, higher lexical density, and much longer average sentence length, demonstrating "dynamic interplay between simplification and complication" (Xiao and Dai 2014, 26).

In the following sections, we will examine each feature in detail, and try to observe the main effects of mediation and genre, as well as their interaction effects on their variation patterns.

	Non-mediated (NS)			Inte	Interpreted (SI)			Translated (WT)		
Mediation subgenre	Q&A_type1 (1)	Q&A_type2 (2)	Debates (3)	1	2	3	1	2	3	
STTR	40.31	40.42	38.40	37.22	35.66	36.42	39.70	36.83	39.79	
List heads	61.37	62.05	64.20	65.16	66.46	66.87	62.26	65.13	62.98	
Lexical density	54.72	56.08	53.16	55.77	55.73	54.11	57.27	58.02	56.37	
ASL	18.48	19.62	21.59	15.48	21.60	15.93	21.78	27.20	23.01	

Table 4. Overall variation patterns of simplification features across sub-genres

^{1.} Due to the limited space of this table, we have temporarily labeled genre *Q&As_type1* as 1, *Q&As_type2* as 2, and *Debates* as 3. Meanwhile, we have dropped the "%" symbol for all percentage values for STTR, list heads, and lexical density also due to the limited space.

4.2.1 Standardized type-token ratio

Standardized type-token ratio shows significant main effects for both mediation (F (2,468)=52.868, p<.001) and genre (F (2,468)=9.705, p<.001), as can be attested in Figure 1. Interpreted language is characterized by much lower STTR than both non-mediated spoken language and translation, meaning that interpreters resort to much less varied vocabularies than native speakers and translators. In terms of genre variation, $Q & As_t y pe2$, in general, seem to be less lexically varied than the other two genres (i.e., $Q & As_t y pe1$ and *Debates*).

A significant interaction effect (F(2, 468) = 9.729, p < .001) between mediation and genre in STTR has been identified, meaning that the effects of mediation status and genre categories are not independent of each other. Pairwise comparisons reveal that there is a significant effect of mediation status on genre variation and vice versa. For Q&As_type1, interpreted language differs significantly from both non-mediated spoken language (p < .001) and translated language (p < .001), with its mean STTR being 37.224 (sd=.367), while the mean for non-mediated language is 40.310 (sd=.374) and for translation 39.697 (sd=.325). For Q&As_type2, interpreting differs significantly from non-mediated spoken language (p < .001), but statistical significance is missing from an intermodal perspective (p=.095), despite the much lower mean value in interpreting as seen in Figure 2. For Debates, there is statistically significant difference from both comparable (p < .001) and intermodal (p < .001) point of view, with the mean STTR for interpreting being 36.419 (sd=.311), while for non-mediated spoken language the value is 38.400 (sd=.338) and for translated language 39.785(sd = .269).

In terms of genre variation within the same level of mediation status, i.e., for non-mediated NS, there is no statistical significance in STTR (p=.862) in $Q & As_type i$ when compared to $Q & As_type2$; however, each of these two shows statistically significant differences in relation to *Debates* (with p < .001 and p=.001, respectively). For simultaneous interpreting, exactly the opposite trends were observed, where only $Q & As_type1$ versus $Q & As_type2$ demonstrate statistically significant difference (p<.05) in STTR, while comparisons between the other two combinations show no significant relation (p=.095 for $Q & As_type1$ versus *Debates*, and p=.202 for $Q & As_type2$ versus *Debates*). For translations, STTR shows distinct variation patterns, where there is no statistical significance between $Q & As_type1$ and *Debates* (p=.834), yet there are for the other combinations (both p<.001).



Figure 1. Distribution of STTR according to mediation status (left) and genre categories (right)



Figure 2. Distribution of STTR in three genre categories across three mediation status

4.2.2 List heads

List heads show significant main effects for both mediation (F(2, 468) = 79.639, p < .001) and genre (F(2, 468) = 27.714, p < .001). As illustrated in Figure 3, list heads in interpreted texts are much higher than in both non-mediated and translated texts, meaning that simultaneous interpreters use more repetitive vocabulary than both native speakers and translators. For genre variation, $Q & As_t ype2$,

in general, are characterized by more repetitive language use. This reflects a similar result to the distribution to the STTR in Figure 1.

Interaction between mediation and genre also shows a significant effect (F (2,468)=9.363, p<.001). Pairwise comparisons for all three genres (see also Figure 4) reveal that texts of different mediation levels differ from each other significantly. $Q & As_t ype_1$ revealed statistically significant distinctions in both comparable (p<.001) and intermodal (p<.001) perspectives for interpreted language. Likewise, *Debates* showed statistical significance between interpreting and non-mediated spoken language (p<.001), and between interpreting and translated language (p<.001). The results for $Q & As_t ype_2$ were far more complex. For example, list heads in interpreting were statistically higher in relation to both non-mediated speech (p<.001) and translation (p<.05), whereas the same feature resulted in 66.465% (sd=.461) for interpreting, while 62.050 (sd=.445) for non-mediated speech and 65.133 (sd=.437) for translation.

However, examination of genre variation within texts of the same level of mediation status shows how the two main effects interact with each other. For non-mediated spoken texts, list heads in $Q & As_type_1$ and $Q & As_type_2$ do not show any statistically significant difference (p=.201). By contrast, such a comparison in interpreted texts and in translated texts reports statistical significance (p<.05 and p<.001, respectively). List heads in $Q & As_type_1$ versus *Debates* in non-mediated and interpreted texts show statistically significant differences as p<.001 and p<.001, respectively, without any statistical significance in translated texts in this regard (p=.06). In terms of list heads in $Q & As_type_2$ versus *Debates*, while non-mediated and translated texts report statistical significance (p<.001 and p<.001, respectively), they do not reach statistical significance in interpreted texts (p=.457).



Figure 3. Distribution of list heads according to mediation status (left) and genre categories (right)



Figure 4. Distribution of list heads in two genre categories across three mediation status

4.3 Information density – Lexical density

In terms of lexical density, there are significant main effects for both mediation (F (2,468)=82.527, p<.001) and genre (F (2,468)=52.545, p<.001). Figure 5 shows that there are noticeable differences among texts of three mediation statuses. Interpreted texts are characterized by higher lexical density than non-mediated texts, with lower lexical density than translated texts. In terms of genre variation, *Debates* are less lexically dense than the two Q&As types, indicating much lower information density of that specific genre.

Interaction between mediation and genre also reveals statistical significance (F(2, 468) = 2.653, p < .05). For $Q & As_type1$, interpreting differs significantly from both non-mediated language (p < .05) and translated language (p < .001), with its mean lexical density (mean = 55.771, sd = .248) being higher than that in the spoken language data (mean = 54.719, sd = .253) while lower than translation (mean = 57.273, sd = .220), conforming to the general trend reported above. For $Q & As_type2$, while intermodal comparison is statistically significant (p < .001), comparable comparison is not (p = .473). For *Debates*, both comparable and intermodal comparison reports statistically significant differences (p < .05 and p < .001 respectively), with interpretations indexing a higher level of informativeness than non-interpretations.

Genre variation with respect to lexical density within the same level of mediation has also been identified, especially in non-mediated texts. For noninterpretations, there is a very strong effect of genre in lexical density since all three genres differ from each other to a statistically significant degree (p<.001 for all). However, in interpreted texts, differences between $Q & As_t ype1$ versus $Q & As_t ype2$ have been leveled out, as their statistical significance is missing (p=.932). Nevertheless, this is not the same case for lexical density in $Q & As_t ype1$ versus *Debates* and $Q & As_t ype2$ versus *Debates*, where statistically significant differences have been observed (p<.001 for both). Similar patterns have also been found in translations, where $Q & As_t ype1$ are homogeneous with $Q & As_t ype2$ in lexical density (p=.056), but not in the remaining sub-genre comparisons.



Figure 5. Distribution of lexical density according to mediation status (left) and genre categories (right)

4.4 Syntactic complexity – Average sentence length

Average sentence length in interpreted texts is significantly shorter than both non-mediated texts and translated texts, showing a strong main effect for mediation status (F(2,468)=138.139, p<.001). Texts of different genre categories also exhibit a main effect (F(2,468)=47.963, p<.001) in average sentence length, with $Q & As_type2$ being the most remarkable in this respect.

The interaction between mediation and genre is statistically significant as well (F(2,468)=15.565, p<.001). For each level of genre categories, the average sentence length in texts of three mediation statuses consistently displays statistical significance. For both $Q & As_t ype1$ and Debates, interpreted texts are characterized by much shorter average sentence lengths compared with non-mediated texts and translated texts. By contrast, the average sentence length in interpreted $Q & As_t ype2$ data is longer than in non-mediated counterparts but shorter than its corresponding translations. However, different patterns have been identified in



Figure 6. Distribution of lexical density in two genre categories across three mediation status

terms of the interaction effect of mediation on genre variation. For non-mediated texts, ASL in $Q & As_t ype1$ and $Q & As_t ype2$ does not show statistical significance (p=.133), but the same pairs of comparison in interpreting and translation yield statistically significant results (both p < .001). $Q & As_t ype1$ versus *Debates* differ significantly in ASL in non-mediated texts and translated texts (p < .001 and p < .05 respectively), but do not for interpreted texts (p=.457). For the comparison between $Q & As_t ype2$ versus *Debates*, ASL differs significantly in all texts across all mediation status (p < .05, p < .001, and p < .001 for non-mediated, interpreted, and translated texts, respectively).

5. Discussion

The above results show that overall, the effect of mediation status (i.e., being non-mediated, interpreted, or translated) is much stronger than the interaction effects between mediation status and genre categories in terms of the variation patterns of simplification features. Viewed from a comparable perspective, interpreted language is characterized by more simplified language use in three out of the four parameters (i.e., STTR, list heads, and average sentence length). That is,



Figure 7. Average sentence length according to mediation status (left) and genre categories (right)



Figure 8. Average sentence length in two genre categories across three mediation status

along the dimensions of lexical diversity and syntactic complexity, interpreting is less lexically varied, more repetitive, and less sophisticated than non-mediated spoken language. In terms of informativeness operationalized as lexical density, interpretations are more lexically dense than non-interpretations. This is contrary to the simplification hypothesis. These findings are in line with previous studies by Kajzer-Wietrzny (2012, 2015) based on the Translation and Interpreting Corpus (TIC) data.

In terms of genre variation, comparison between interpretations and noninterpretations confirms the simplification hypothesis in general, except for lexical density and average sentence length in Q&As_type2. Lexical density in interpreted Q&As_type2 is slightly lower than their non-interpreted counterparts, but no statistical significance has been reported. The average sentence length in interpreted Q&As_type2 is much higher than their non-interpreted counterparts. This is an interesting finding because it is contradictory to the general expectation of a reduction of sentence complexity in interpreting. This deviation from the overall trend of simplification in interpreted language shows the special status of genre Q&As_type2. A closer examination reveals that this specific "deviation" of lexical patterns in Q&As_type2 may be attributed to the mode of delivery rather than genre per se. As briefly introduced in Section 3, in distinction from the other two genre categories, Q&As_type2 contains both scripted, written-to-beread Q&As as well as prepared but unscripted Q&As, leading to a mixed combination of features of both orality and literacy. In comparison, the other two genres contain few scripted speeches. Kajzer-Wietrzny (2015) observes an influence of the mode of delivery on the degree of repetitiveness but not on lexical density. Her study shows that interpretations of speeches that were originally spoken were more repetitive than non-interpretations. However, this is not supported in our study, as the interpreted data in the Q&As_type2 sub-corpus, which contains a large proportion of written-to-be-read speeches, turns out to be the most repetitive sub-genre category with the lowest STTR and highest list head coverage (see Figures 2 and 4).

In general terms, the contradictory status of lower lexical diversity versus higher information density in interpretations versus non-interpretations has been confirmed. Moreover, this has also been identified for the Cantonese and English language pair, conforming to previous scholarship based on other language combinations. Instead of offering further evidence of the mixed patterns of lexical simplification in interpreting, we argue this may suggest that "translational simplification is not a pure, simple phenomenon" (Xiao and Dai 2014, 26). In keeping with the complexity of many other translational phenomena, the seemingly contradictory linguistic realizations in interpreting are the product of an interplay of factors. Shlesinger (1995) proposes three constraints intrinsic to simultaneous interpreting: the time constraint, the linearity constraint, and the (un)shared knowledge constraint. The first two of these can offer some explanations for the observed simplification patterns (higher repetitiveness and lower syntactic complexity) in the current study. The time constraint refers to the fact that simultaneous interpreters, unlike native speakers of English, are paced by original speakers and have no control over this dimension. Therefore, to avoid cognitive saturation, they have to perform some trade-offs during the multitasking efforts of listening and comprehending, memory recall, production, and coordination (Gile 2009). This can be achieved, in part, through strategies such as relying on high-frequency words instead of searching for low-frequency and more varied words, which leads to more repetitive language use when they are heavily taxed by a syntactic processing operation (see Gile 2009; Shlesinger 1995). The linearity constraint means that the text/speech becomes available gradually so that the interpreters "are forced to focus on relatively short units of language" (Shlesinger 1995, 194), which may have explained the much shorter average sentence length in interpreted speech. The exception of Q&As_type2, as discussed above, might be attributed to the mode of delivery of both source speakers and interpreters, since for the scripted speeches, interpreters are also often prepared with texts beforehand, and therefore they may feel less constrained in this respect. Another possible reason for this (relatively consistent) higher repetitiveness and lower syntactic complexity in interpreting may also be related to the working direction of our corpus data. The interpreters have done so from their A language (Cantonese) into their B language (English), and second language processing is believed to be more cognitively challenging. Dayter (2018) also observes a possible influence of working direction on the contradictory lexical patterns in interpreted English from Russian and interpreted Russian from English with respect to simplification and explication (reviewed in Section 2). However, in contrast to ours, she argues that interpreting into the B language (one-third of the Ru>En samples) may lead to certain behaviors such as over-correction, which eventually gives rise to "a high content word ratio and an elaborate, varied vocabulary" (ibid., 257). We argue quite the opposite since working into the B language adds a further layer of cognitive constraint to simultaneous interpreters and reveals more prominent simplification patterns in our data.

Moving on to an intermodal analysis, interpreted language is found to be more simplified than translated language along all three dimensions (i.e., lexical diversity, informativeness, and syntactic complexity). To be specific, interpreting is characterized by significantly lower STTR, higher list head coverage, lower lexical density, and shorter average length than translation, indicating a strong influence of modality or modes of mediation.

For genre variation, certain genres do not show statistically significant differences in STTR and list heads in interpretations as opposed to the translations. $Q & As_type2$, for example, did not show statistical significance in STTR variation from an intermodal perspective, although STTR in interpreted $Q & As_type2$ is lower than its translated counterpart. Nevertheless, the intermodal analysis overall lends support to previous findings that interpreters simplify more than translators (e.g., Bernardini et al. 2016; Ferraresi et al. 2018). This also offers further evidence to the claim that interpreting is "an extreme case of translation" (Shlesinger and Ordan 2012, 54), where the typical features of translations become all the more prominent in interpretations. As far as the reasons are concerned, we believe that both the ontological differences between translation and interpreting as intrinsically pieces of written and spoken discourses and the inherent differences between translation and interpreting as two modes of mediation contribute to this simplification result. Previous studies (Chafe 1982; Gibson et al. 1966; Redekker 1984) have offered ample evidence for the more informal, fragmented, and simplified nature of spoken language, while written language is found to be more formal, integrated, and elaborate. However, we also acknowledge the fact that it is very difficult to disentangle the two (ontological differences versus different modes of mediation), which requires a well-designed study in the future.

In addition to the general findings about the simplification hypothesis, we also find limited support for the "leveling out" (Baker 1993) or "convergence" hypothesis (Laviosa 2002). Baker's (1996, 184) support for the leveling out hypothesis rests on her observation that "the individual texts in an English translation corpus are more like each other in terms such as lexical density, type-token ratio and mean sentence length than the individual texts in a comparable corpus of original English." This implies reduced variation between oral and written translations (Shlesinger 1989). Based on the visualizations of the four simplification features within genre categories, we have found that genre variations within interpretations are not homogeneous (if not less homogeneous than noninterpretations) since variations consistently exist with respect to the four simplification features. Therefore, this study offers counterevidence to the "leveling out" hypothesis. This "heterogeneity" is manifested especially in Q&As_type2, which may again suggest that mode of delivery may be a more influencing factor than genre per se. In future studies, more soundly designed studies need to be carried out to isolate the possible influence of different modes of delivery and genre categories.

6. Conclusion

The present study has attempted to examine genre variation on simplification patterns of the interpreted outputs from both comparable and intermodal perspectives based on the Cantonese-to-English language pair. The comparable analysis confirms the more repetitive and syntactically less complex nature of interpreted language, but it offers counterevidence to simplification with respect to lexical density. These patterns are generally consistent across genre comparisons, except for lexical density in the Q&As_type2 interpreting data. The intermodal comparison demonstrates the more repetitive and simplified nature of

interpreting versus translation, irrespective of genre differences. This indicates a strong influence of modality or modes of mediation on simplification.

Despite the generally consistent simplification patterns of interpreted language from both comparable and intermodal perspectives, the statistically significant interaction effects between mediation and genre, as well as the main effect of genre, indicate that genre variations do exist, regardless of the mediation status, lending limited support to the leveling out hypothesis (Baker 1993). These findings indicate that more fine-grained analysis on genre or register variation and features of interpreted language needs to be carried out since the linguistic manifestations of translation and interpreting, as also argued by Kruger and Van Rooy (2012, 62), "are variable and subject to the influence of a variety of factors, amongst others register."

To answer Shlesinger and Ordan's (2012, 44) question concerning "whether interpreting is essentially 'the same as' translation, other than the fact that happens to be oral; whether it is first and foremost a form of speech, with distinct spoken-like features that override its translational ontology," we conclude, based on our LegCo+ data and the variation patterns of four simplification features, that interpreting can be regarded as an extreme case of both spoken and mediated discourse.

One shortage of the current study lies in the loose categorization of genres due to the limited genre types in the legislative settings. However, we argue that our major goal is to arouse discussions on possible genre variation in features of interpreted language when contradictory findings are reported. Our study, based on a loose categorization of genres, has illustrated that there are indeed interaction effects between mediation status and genre, meaning that the differences among texts of different mediation statuses may depend upon genre categories and vice versa. Focusing only on the general picture of ontological differences will inevitably mask the nuances of differences within, and thus more fine-grained analysis should be carried out to avoid this pitfall.

Finally, as rightfully expressed by Xiao and Dai (2014, 26), "translational simplification is not a pure, simple phenomenon," and many other researchers agree that simplification cannot be a simple matter for either translation or interpreting (e.g., Evert and Neumann 2017; De Sutter and Lefer 2020). Our final goal is not simply to confirm nor refute the proposed hypotheses but rather, through the analysis of surface manifestations of these mediated products, to get a better understanding of the possible cognitive, social and cultural constraints which contribute to the final linguistic realizations. More sophisticated techniques should be employed in future works to unveil their "multidimensional and multifaceted nature" (De Sutter and Lefer 2020) since translation and interpreting are

far from simple procedures and genre types as well as other possible factors layer these with complexity.

Funding

The work was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (PolyU/RGC 15602621), and partially supported by the Fundamental Research Funds for Central Universities (Grant ID: 2022JS004).

References

Baker, Mona. 1993. "Corpus Linguistics and Translation Studies: Implications and Applications." In *Text and Technology: In Honour of John Sinclair*, edited by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 233–250. Amsterdam and Philadelphia: John Benjamins Publishing Company.

https://doi.org/10.1075/z.64.15bak

- Baker, Mona. 1996. "Corpus-based Translation Studies: The Challenges That Lie Ahead." Benjamins Translation Library 18: 175–186. https://doi.org/10.1075/btl.18.17bak
- Bernardini, Silvia, Adriano Ferraresi, and Maja Miličević. 2016. "From EPIC to EPTIC Exploring Simplification in Interpreting and Translation from an Intermodal Perspective." *Target* 28 (1): 61–86.

https://doi.org/10.1075/target.28.1.03ber

Biber, Douglas. 1988. Variation Across Speech and Writing. Cambridge: Cambridge University Press.

https://doi.org/10.1017/CBO9780511621024

- Blum-Kulka, Shoshana, and Eddie. A. Levenston. 1978. "Universals of Lexical Simplification." Language Learning 28 (2): 399–415.
 - https://doi.org/10.1111/j.1467-1770.1978.tb00143.x
- Chafe, Wallace. 1982. "Integration and Involvement in Speaking, Writing, and Oral Literature." In *Spoken and Written Language: Exploring Orality and Literacy*, edited by Deborah Tannen, 35–54. Norwood, NJ: ABLEX Pub. Corp.
- Chen, Jiansheng 陳建生, and Cui Yani 崔亞妮. 2010. "Jiyu yuliaoku de zhongguo zhenfu gongzuobaogao yingyiben cihui tezheng yanjiu" 基於語料庫的中國《政府工作報告》 英譯本詞匯特征研究 [A Corpus-based study on lexical features in the English translation of *Report on the Work of the Government*]. *Dangdai waiyu yanjiu* 當代外語研 究 [Contemporary foreign languages studies] 6: 39–43.
- Dayter, Daria. 2018. "Describing Lexical Patterns in Simultaneously Interpreted Discourse in a Parallel Aligned Corpus of Russian-English Interpreting (SIREN)." *Forum* 16 (2): 241-264.

https://doi.org/10.1075/forum.17004.day

De Sutter, Gert, and Marie-Aude Lefer. 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. https://doi.org/10.1080/0907676X.2019.1611891

- Evert, Stephen, and S. Neumann. 2017. "The Impact of Translation Direction on Characteristics of Translated Texts. A Multivariate Analysis for English and German." In *Empirical Translation Studies: New Theoretical and Methodological Traditions*, edited by Gert De Sutter, Isabelle Delaere, and M.A. Lefer, 47–80. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110459586-003
- Ferraresi, Adriano, Silvia Bernardini, Maja Petrović, and Marie-Aude Lefer. 2018. "Simplified or Not Simplified? The Different Guises of Mediated English at the European Parliament." *Meta* 63 (3): 717–738. https://doi.org/10.7202/1060170ar
- Gibson, James W., Charles R. Gruner, Robert J. Kibler, and Francis J. Kelly. 1966. "A Quantitative Examination of Differences and Similarities in Written and Spoken Messages." *Speech Monographs* 33 (4): 444–451. https://doi.org/10.1080/03637756609375510
- Gile, Daniel. 2009. *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam and Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/btl.8
- Hu, Kaibao 胡開寶, and Qing Tao 陶慶. 2010. "Hanying huiyi kouyi yuliaoku de chuangjian yu yingyong yanjiu" 漢英會議口譯語料庫的創建與應用研究 [The compilation and use of the Chinese-English conference interpreting corpus]. *Zhongguo fanyi* 中國翻譯 [Chinese translators journal] 5: 49–56.
- Hu, Xianyao, Richard Xiao, and Andrew Hardie. 2016. "How do English Translations Differ from Non-translated English Writings? A Multi-feature Statistical Model for Linguistic Variation Analysis." *Corpus Linguistics and Linguistic Theory* 15 (2): 347–382. https://doi.org/10.1515/cllt-2014-0047
- Jantunen, Jarmo Harri. 2001. "Synonymity and Lexical Simplification in Translations: A Corpus-based Approach." *Across Languages and Cultures* 2 (1): 97–112. https://doi.org/10.1556/Acr.2.2001.1.7
- Kajzer-Wietrzny, Marta. 2012. "Interpreting Universals and Interpreting Style." Ph.D. diss. Adama Mickiewicza University.
- Kajzer-Wietrzny, Marta. 2015. "Simplification in Interpreting and Translation." *Across Languages and Cultures* 16 (2): 233–255. https://doi.org/10.1556/084.2015.16.2.5
- Kruger, Haidee, and Bertus Van Rooy. 2012. "Register and the Features of Translated Language." *Across Languages and Cultures* 13 (1): 33–65. https://doi.org/10.1556/Acr.13.2012.1.3
- Kruger, Haidee, and Bertus Van Rooy. 2018. "Register Variation in Written Contact Varieties of English: A Multidimensional Analysis." *English World-Wide* 39 (2): 214–242. https://doi.org/10.1075/eww.00011.kru
- Laviosa, Sara. 1997. "How Comparable can 'Comparable Corpora' be?" *Target* 9 (2): 287–317. https://doi.org/10.1075/target.9.2.05lav
- Laviosa, Sara. 1998. "Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose." *Meta* 43 (4): 557–570. https://doi.org/10.7202/003425ar

Laviosa, Sara. 2002. Corpus-based Translation Studies: Theory, Findings, Applications (Vol. 17). Amsterdam: Rodopi.

https://doi.org/10.1163/9789004485907

- Li, Dechao 李德超, and Kefei Wang 王克非. 2012. "Hanying tongchuan zhong cihui moshi de yuliaoku kaocha" 漢英同傳中詞匯模式的語料庫考察 [A corpus-based study on lexical patterns in simultaneous interpreting from Chinese into English]. *Xiandai waiyu* 現代外 語 [Modern foreign languages] 35 (4): 409-415.
- Lü, Qianxi, and Junying Liang. 2018. "Is Consecutive Interpreting Easier than Simultaneous Interpreting? – A Corpus-based Study of Lexical Simplification in Interpretation." *Perspectives* 27 (1): 91–106.

https://doi.org/10.1080/0907676X.2018.1498531

- Mauranen, Anna. 2000. "Strange Strings in Translated Language: A Study on Corpora." In Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects, edited by Maeve Olohan, 119–141. London: Routledge.
- Qin, Hongwu 秦洪武, and Kefei Wang 王克非. 2009. "Jiyu duiying yuliaoku de yingyi hanyu yuyan tezheng fenxi" 基於對應語料庫的英譯漢語言特征分析 [A parallel corpus-based study of Chinese as target language in EC translation]. *Waiyu jiaoxue yu yanjiu* 外語教學 與研究 [Foreign language teaching and research] 2: 131–136.
- Redeker, Gisela. 1984. "On Differences Between Spoken and Written Language." *Discourse Processes* 7 (1): 43–55.

https://doi.org/10.1080/01638538409544580

- Russo, Mariachiara, Claudio Bendazzoli, and Annalisa Sandrelli. 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. https://doi.org/10.1075/forum.4.1.10rus
- Sandrelli, Annalisa, and Claudio Bendazzoli. 2005. "Lexical Patterns in Simultaneous Interpreting: A Preliminary Investigation of EPIC (European Parliament Interpreting Corpus)." In *Proceedings from the Corpus Linguistics Conference*, 1–18. Birmingham: University of Birmingham.
- Sandrelli, Annalisa, Claudio Bendazzoli, and Mariachiara Russo. 2010. "European Parliament Interpreting Corpus (EPIC): Methodological Issues and Preliminary Results on Lexical Patterns in Simultaneous Interpreting." *International Journal of Translation* 22: 165–203.
- Schmid, Helmut. 1994. TreeTagger A Language Independent Part-of-Speech Tagger. http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/
- Scott, Michael. 2012. WordSmith Tools (Version 6) [Computer software]. Lexical Analysis Software.

https://lexically.net/wordsmith/downloads/

- Shlesinger, Miriam. 1989. "Simultaneous Interpretation as a Factor in Effecting Shifts in the Position of Texts on the Oral-Literate Continuum." Ph.D. diss., Tel Aviv University.
- Shlesinger, Miriam. 1995. "Shifts in Cohesion in Simultaneous Interpreting." *The Translator* 1 (2): 193–214.

https://doi.org/10.1080/13556509.1995.10798957

Shlesinger, Miriam. 2008. "Towards a Definition of Interpretese: An Intermodal, Corpus-based Study." In *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*, edited by Hansen, Gyde, Andrew Chesterman, and Heidrun G. Arbogast, 237–253. Amsterdam and Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/btl.80.18shl Shlesinger, Miriam, and Noam Ordan. 2012. "More Spoken or More Translated?: Exploring a Known Unknown of Simultaneous Interpreting." *Target* 24 (1): 43–60. https://doi.org/10.1075/target.24.1.04shl

- Williams, Donna A. 2005. "Recurrent Features of Translation in Canada: A Corpus-based Study." PhD diss., University of Ottawa.
- Xiao, Richard. 2010. "How Different is Translated Chinese from Native Chinese?: A Corpusbased Study of Translation Universals." *International Journal of Corpus Linguistics* 15 (1): 5–35.

https://doi.org/10.1075/ijcl.15.1.01xia

- Xiao, Richard, and Ming Yue. 2009. "Using Corpora in Translation Studies: The State of the Art." In *Contemporary Corpus Linguistics*, edited by Paul Baker, 237–262. London: Continuum.
- Xiao, Richard, and Guangrong Dai. 2014. "Lexical and Grammatical Properties of Translational Chinese: Translation Universal Hypotheses Reevaluated from the Chinese Perspective." *Corpus Linguistics and Linguistic Theory* 10 (1): 11–55. https://doi.org/10.1515/cllt-2013-0016
- Xiao, Richard, and Xianyao Hu. 2015. "The Features of Translational Chinese and Translation Universals." In *Corpus-based Studies of Translational Chinese in English-Chinese Translation*, edited by Richard Xiao, and Xianyao Hu, 157–167. Berlin: Springer. https://doi.org/10.1007/978-3-642-41363-6_8

Résumé

Dans cet article, nous explorons la variation et la simplification des genres dans le langage interprété d'un point de vue à la fois comparable (interprété vs non interprété/non médiatisé) et intermodal (interprété vs traduit). Nous nous appuyons sur un corpus, nouvellement construit, comparable dans une seule direction et intermodal nommé LegCo+, un corpus présentant les démarches législatives du Conseil législatif de Hong Kong et leurs traductions et interprétations du cantonais vers l'anglais, ainsi que des discours pléniers originaux prononcés par des natifs anglophones au Parlement britannique. Suivant les traditions de recherche antérieures, nous étudions les modèles de variation de quatre caractéristiques de simplification en trois dimensions, à savoir le ratio type/token standardisé (STTR) et les têtes de liste pour la diversité lexicale, la densité lexicale pour l'informativité, et la longueur moyenne des phrases pour la sophistication lexicale. Nous visons à découvrir les effets de la médiation et du genre, ainsi que leurs effets d'interaction sur la variation linguistique. Nos résultats indiquent que les textes de différents statuts de médiation et de différentes catégories de genre varient en fonction des modèles de simplification. Dans une perspective compararée, les interprétations reposent sur une gamme de vocabulaires plus étroite que les non-interprétations, mais elles sont également plus informatives, et cette informativité dépend des catégories de genre. D'un point de vue intermodal, les interprétations montrent des modèles cohérents, étant plus simplifiés que les traductions, ce qui indique un fort effet de modalité (ou mode de médiation).

Mots-clés : variation de genre, simplification, comparable, intermodal, modèles de variation