Article

CADERNOS de TRADUÇÃO

Arabic sentence patterns in interpreted, translated, and original speeches: A corpus-based approach

Aladdin Al Zahran
Sohar University
Sohar, Sultanate of Oman
aalzahran@su.edu.om
https://orcid.org/0000-0001-5382-6007

Rafik Jamoussi
Sohar University
Sohar, Sultanate of Oman
rjamoussi@su.edu.om
https://orcid.org/0000-0003-3036-8968

Eassa Ali
Sohar University
Sohar, Sultanate of Oman
eali@su.edu.om
https://orcid.org/0009-0002-8271-1545

Abstract: In simultaneous interpreting (SI) from a subject-verb-object (SVO) language such as English, Arabic simultaneous interpreters are confronted with the dilemma of choosing between the default, dominant, and unmarked V-initial or VSO structure, or a derived and marked S-initial or SV(O) structure. The choice of the former may cause memory overload and omissions or failure due to prolonged time lag and excessive storage of input chunks, especially if structural asymmetry coincides with such extreme factors as high source language (SL) input rate, structural complexity and/or information density. By contrast, the choice of any of the SV(O) structures available in Arabic can help interpreters closely follow the structure of SL discourse and avoid potential risks despite the shift of focus and emphasis brought about by the marked SV(O) structure. Using a corpus-based analysis of an English-to-Arabic parallel multimodal corpus of mediated and non-mediated discourse, the article seeks to determine which structure is used more frequently in simultaneously interpreted speeches compared to translated and original ones. The results indicated the dominance of the S-initial structure in SI, whereas the V-initial structure dominated translations and original speeches and that the choice of structure was impacted by mediation type (SI *vs.* translation) and language modality (written *vs.* spoken). These findings help relate the choice of structure to the SI process and suggest that the interpreters adopted structure-based processing. The results contribute to the research on a relatively under-investigated language pair and are valuable for informing SI training.
Keywords: language-pair specificity; interpreting strategies; syntactic asymmetry; SVO and VSO structures; English-Arabic simultaneous interpreting.

CT
Cadernos de Tradução, 45, 2025, e104911
Graduate Program in Translation Studies
Federal University of Santa Catarina, Brazil. ISSN 2175-7968
DOI https://doi.org/10.5007/2175-7968.2025.e104911

1. Introduction

In simultaneous interpreting (SI), structural asymmetries between language pairs have been extensively discussed within the topic of "language-pair specific factors" (Wilss, 1978, p. 350) and "language-pair specificity" (Setton, 1999, p. 55), which generally fall under the wider subject area of contrastive interpreting studies. Language-pair specific factors, whether linguistic or cultural (Setton, 1993), have been viewed as problem-triggers (Gile, 2009; Shamy & de Pedro Ricoy, 2017; Zou & Wang, 2023). The discussions reported in the literature have focused specifically on the challenges of language-pair specific factors or ease of processing due to similarities between the structures of language pairs. These studies have generally explored the impact of structural asymmetries on the process and, to a larger extent, product of SI. Whenever difficulties were identified, discussions have consistently explored the strategies or tactics adopted by simultaneous interpreters to cope with such challenges. In this article, we use the term *coping strategies* in line with established interpreting literature, where they are referred to as "coping tactics" (Gile, 2009, p. 191), "SI skills", or "acquired strategies" (Setton, 1999, p. 50). They are understood as recurrent, observable behavioural patterns employed to cope with cognitive overload or extreme conditions and, if properly introduced, can be acquired and progressively automatized (Kirchhoff, 2002; Riccardi, 1996). See also Dayter (2020) for a more detailed discussion of strategies in interpreting.

Many language pairs have been involved in the examination of structural asymmetry in SI. This is especially true when one of the working languages in the language pair has a left branching or verb-last (subject-object-verb, SOV) structure or is characterized by high frequency of multiple and complex embeddings, topicality, and long and complex and/or semantically dense sentences or parts thereof (Al Zahran, 2021; El-Zawawy, 2021; Kade & Cartellieri, 1971; Meuleman & Besien, 2009; Setton, 1993).

The present work aims primarily to report on a corpus-based study of Arabic sentence patterns in simultaneously interpreted, translated and original speeches. This research integrates four sub-corpora that comprise authentic English conference speeches, their Arabic SIs and written translations, as well as original Arabic speeches. By combining these distinct linguistic datasets, our research seeks to uncover insights into the dynamics of Arabic discourse whether simultaneously interpreted, translated, or originally produced by determining the structural patterns mostly used across these language modalities. In addition to relying on a quadruple-corpus design, the study's contribution consists in applying and refining established models such as Gile's (1999, p. 153) "tightrope hypothesis" and Setton's (1999) cognitive–pragmatic approach in the underexplored English–Arabic context, thereby extending their relevance to a new typological setting. We thus anticipate that the findings of our study can make a significant contribution to the existing body of literature on contrastive interpreting studies and particularly language-pair specificity in SI, as well as a meaningful contribution to the professional practice and pedagogy (Rodríguez-Inés & Gallego-Hernández, 2016) of (Arabic) SI. The present article aims to examine the issue of language-pair specificity from a syntactic perspective, seeking to determine the structure used more frequently in a parallel multimodal corpus comprising three sub-corpora, namely Arabic SIs of English conference speeches, Arabic translations of the same English speeches, and original Arabic conference speeches.

We expect the study to offer an insight into a choice between the default, dominant, and unmarked verb-subject-object (VSO) structure (Dahlgren, 2009; Fassi-Fehri, 1993; Hoyt, 2009; Mohammad, 2000; see also Aoun et al., 1994) on the one hand, and a derived (Abdul-Raof, 1998), marked subject-verb-(object) (SV(O)) structure on the other.

In Arabic SI, the choice of the default, unmarked VSO structure entails a substantial risk as it may force the interpreter to lag far behind the original speaker and retain relatively large amounts of information in short-term memory (STM), potentially causing excessive cognitive overload, possible loss of information, and potential failure. This risk may increase considerably if structural dissimilarity coincides with other complicating factors or "extreme speech conditions" (Meuleman & Besien, 2009, p. 20), such as high source language (SL) presentation rate, structural complexity (Chmiel et al., 2024; El-Zawawy, 2021), especially in the subject position (Al Zahran, 2021), and/or semantic or information density (Gile, 2011; Kirchhoff, 2002; Setton, 1999). By comparison, the use of the derived, marked SV(O) structure has the advantage of allowing the interpreter to closely follow the identical English SVO structure, thus eliminating the need for prolonged time lag or storage of excessive amounts of information in STM, thereby avoiding all the potential negative consequences. Nonetheless, the use of the marked SV(O) structure has the downside of bringing about a discourse shift in terms of focus and emphasis. For detailed discussions of the dichotomy between Arabic VSO and SV(O) structures, along with authentic examples from the corpus, see Al Zahran (2021) and Al Zahran and Jamoussi (2022).

We can thus hypothesize that (1) the Arabic simultaneous interpreters in our corpus will mostly opt for the S-initial or SV(O) structure because it is easier to process given the constraints of the SI process; (2) the Arabic translators, not being under the same pressure as simultaneous interpreters, will most frequently use the dominant and default V-initial or VSO structure; and (3) the original Arabic speakers will use the V-initial or VSO structure, it being the default, unmarked and dominant structure in Arabic discourse.

1.1 Research questions

Our research questions can therefore be formulated as follows:

> Research question (1): How does the SVO/VSO ratio in Arabic simultaneously interpreted speeches compare with the ratio in the formal (written) translations of these speeches?
>
> Research question (2): How does the SVO/VSO ratio in Arabic simultaneously interpreted speeches compare with the ratio in a comparable spoken corpus of non-mediated original Arabic speeches?

Answers to these questions are provided in the Results section (Section 3) following a survey of the literature on language-pair specificity in the following subsection and a presentation of our research methodology in Section 2.

1.2 Overview of the literature on language-pair specificity

While it has been argued that the impact of language-pair specificity is more conspicuous on SI between language pairs involving Asian than European languages (Setton, 1993), the earliest studies reported in the literature that tackled this issue as early as the 1970s have discussed the challenges in SI between German, with its SOV subordinate and numerous embedded clauses, and other European languages such as English and French (Goldman-Eisler, 1972; Kade & Cartellieri 1971; Kirchhoff, 2002; Moser, 1978; Wilss, 1978). Since then, there has been a notable growth of studies tackling various phenomena of language-pair specificity involving European and non-European, especially Asian (e.g. He et al., 2016; Wang & Gu, 2016), languages and language pairs, employing experimental and empirical research, including retrospective analyses through think aloud protocols (e.g. Chang & Schallert, 2007) and, more recently, corpus-based studies (e.g. Al Zahran, 2021; Al Zahran & Jamoussi, 2022; Dayter, 2020). As mentioned in the introduction, such studies have generally reported on various language-pair-specific problem triggers and some suggested strategies and coping tactics to overcome such issues. Naturally, most of such studies also discuss the pedagogical implications of their findings and the need to integrate language-pair-specific factors and interpreting strategies into the interpreting curriculum.

One of the most frequently cited language-pair-specific factors is word-order or structural mismatches, which have been found to be a cause of various considerable challenges in SI between many language pairs. Too long and/or complex SL constructions were viewed as a cause of longer time lag, use of anticipation, and insertion of "redundant parts" into the target language (TL) rendition (Kade & Cartellieri, 1971, p. 16). The German verb-last structure caused the interpreters in an experiment to store more input chunks and lag longer in German>English SI than in SI from English or French and rely on segmentation based on syntactic units (Goldman-Eisler, 1972). In another experimental study, the rendition of verb-final constructions and constructions with word-order asymmetries in German>Italian SI led to the use of generic verbs by students and restructuring by interpreters (Riccardi, 1996). Interpreting dissimilar structures caused more cognitive load than similar ones in an experimental study of German>English SI using a measure of "task-evoked pupillary responses" (Seeber & Kerzel, 2011, p. 228). Evidence of syntactic segmentations was found in a corpus-based analysis of Russian<>English SI (Dayter, 2020).

The need for prolonged lag and storage of large amounts of data and subsequent retrieval due to language-pair-specific factors and associated difficulties was described in terms of an increase in "processing capacity" requirements (Gile, 1991, p. 19) due to working under the "tightrope hypothesis" (Gile, 1999, p. 153) close to "cognitive saturation" (Gile, 2011, p. 201). It has thus been hypothesized that the level of difficulty or ease in SI partially depends on the specific language pairs involved (Gile, 1991; Wilss, 1978). Moreover, conflicting structures are viewed as "a source of significant additional cognitive load" and though do not per se "rule out SI" (Setton, 2005, p. 71), in fact not only caused challenges but also obstructed SI when coincided with such other variables as semantic density or lack of cognitive context in Chinese>English and German>English SI (Setton, 1999).

Studies on syntactic asymmetry and language-pair-specific factors involving Asian languages mainly concerned languages with "a significant amount of left-branching" (Setton, 1999, p. 53),

including Chinese, Japanese and Arabic. Disparities in word order in Chinese>English SI led to the interpreters' heavy reliance on re-structuring, segmentation, waiting, and anticipation (Dawrant, 1996). Evidence of language-pair specificity was found in Chinese>English SI as a result of differing "rhetorical patterns" in a study combining experimental analysis and retrospective interviews (Chang & Schallert, 2007, p. 172). A tendency to perform reordering of Chinese "front-loaded" structures into back-loaded English ones in 75% of the cases was reported in an experimental study on professional Chinese>English SI (Guo, 2011, as cited in Wang & Zou, 2018, p. 67). Difficulties due to structural disparity in Chinese>English SI forced the interpreters to cope using frequent and exceptionally long pauses and segmentation, often at a high risk of cognitive overload, omissions and inaccuracies (Wang & Gu, 2016).

The late occurrence of Japanese verbs and modified head nouns in Japanese>English led to the frequent use of segmentation and intentional passivation to reduce time lag, as well as to generalizations and omissions due to memory limitations (He et al., 2016).

In an experimental study of English>Arabic SI, errors and omissions were caused by "problematic linear arrangements" (Al-Rubai'i, 2004, p. 249), forcing the interpreters to cope using "Trackings", that is, syntactic tools that allowed them to follow the English structures closely to avoid restructuring and prolonged time lag. A similar strategy, linearization, was found to have been predominantly applied in three Arabic SI versions of the final presidential debate between Hilary Clinton and Donald Trump to cope with various types of complex English structures (El-Zawawy, 2021). Likewise, "form-based" (Isham, 1994) processing was forced upon the interpreters, and cognitive overload, omissions and communication failure occurred due to the coincidence of SVO-VSO syntactic asymmetry and English long or complex initial subjects in a corpus-based analysis of authentic English conference speeches and their Arabic SIs (Al Zahran, 2021). In a study of directionality and competence in English<>Arabic liaison interpreting involving beginner and advanced undergraduates, the advanced students faced no syntactic, grammatical or lexical problems in the English-to-Arabic direction but were challenged at the lexical and grammatical levels in the opposite direction (Al-Jarf, 2022). This finding provides clear evidence of the interplay between language-pair specificity and interpreting direction.

Syntactic asymmetry did not only cause difficulties in interpreting from but also *into* SOV languages. The SI of Obama's 2009 inaugural address from English into Japanese was more challenging than English>French and English>German SI (Gile, 2011). Dutch and German interpreters used extraposition of elements in subordinate clauses to a post-verbal position from the middle field to cope with structural asymmetry in a study of a corpus of interpreted and non-mediated speeches from the European Parliament (Collard et al., 2018).

In the same way language-pair-specific factors caused difficulties and challenges to SI, they were occasionally cited as a source of "ease" of SI when similar structures existed in some language pairs. Thus, in a corpus analytical study of SI from Japanese, the ability to predict sentence endings imposed less processing capacity requirements by limiting concurrent speaking and listening compared to German, French, and English, leading to the assumption that these endings may be considered as a language-specific factor (Gile, 1992, 2009). In a study of "deverbalization" in the recall of the most recent sentence, the need for the "meaning-based" approach increased with greater syntactic asymmetries between language pairs, while syntactic similarities called more for a

"form-based" approach (Isham, 1994, p. 205-208). Morphosyntactical asymmetries between Spanish and Italian necessitated a "more form-based" approach and "processing ease" due to a certain level of similarity, but not identity, between the syntactical and phonological structures. Conversely, asymmetries affecting word form and word order, which could result in ambiguous meanings in Italian despite semantic and phonological similarity to Italian words required a "more meaning-based" approach and "processing depth", underscoring the complexity of interpreting between cognate languages (Russo, 1997, p. 269-270). An experimental study of SI-with-text between English and French found no support for syntactic-semantic restructuring (Setton & Motta, 2007), indicating that interpreters tend to use reordering less often whenever syntactic structures are similar. In an experimental study of SI from English>Spanish, both being SVO languages, the Spanish TL interpretations were compared with Spanish originals and were found to be partially similar to the lexical and grammatical features of the SL texts, leading to the conclusion that SI is largely a "language-specific" task (Alonso Bacigalupe, 2010, p. 50). Syntactic linearity was found to be a highly effective strategy in the Chinese SI of Obama's 2009 Q&A session with Chinese students allowing the interpreters to reduce time lag and risk of cognitive overload by preserving the SL thematic structures in the TL (Chen et al., 2015). A corpus-based analysis of English>Arabic SI revealed that the interpreters used the marked SV(O) structure more frequently than the default or basic VSO structure due to the former's similarity with English SVO structures to avoid prolonged time lag, storage of large input chunks, and potentially ensuing information loss and complete communication failure (Al Zahran & Jamoussi, 2022).

Not only were syntactic asymmetries considered as language-pair specific, but also the strategies and coping tactics used to overcome them were found to be language-pair specific. Prediction was considered as a language-specific factor depending on whether the verb occurs early in the sentence, as in English, or at a later stage, as in German, leading to the assumption that prediction is "easier" and "much more reliable" in English>German SI than vice versa (Moser, 1978, p. 360). In a discussion of syntactic anticipation in German>English by Wilss (1978, p. 350), the SI process was described as "language-pair specific" as it is governed by syntactic asymmetry. Time lag was found to be language-pair specific as it is "*affected by the language combination*" by Seleskovitch and Lederer (1995, p. 131, our emphasis), who traditionally opposed linguistic differences and language-pair-specific factors in favor of sense-based processing. Anticipation was found to be language-pair specific in corpus-analytical studies of German>French (Van Besien, 1999) and German>Greek (Liontou, 2011) SI. Language-pair specificity was experimentally demonstrated when the interpreters were forced to apply anticipation and reformulation more frequently and lag longer in German>Italian than Dutch>Italian SI (Bevilacqua, 2009).

Having surveyed the literature on language-pair specificity, our corpus-based research addresses a notable gap in the current body of knowledge: the lack of studies combining analyses of SIs and written translations of authentic conference speeches and comparable original speeches. This novel multimodal approach offers a more comprehensive view of syntactic patterns in English>Arabic SI and aims to advance contrastive interpreting studies, with direct implications for interpreter training and professional practice.

2. Materials and methods

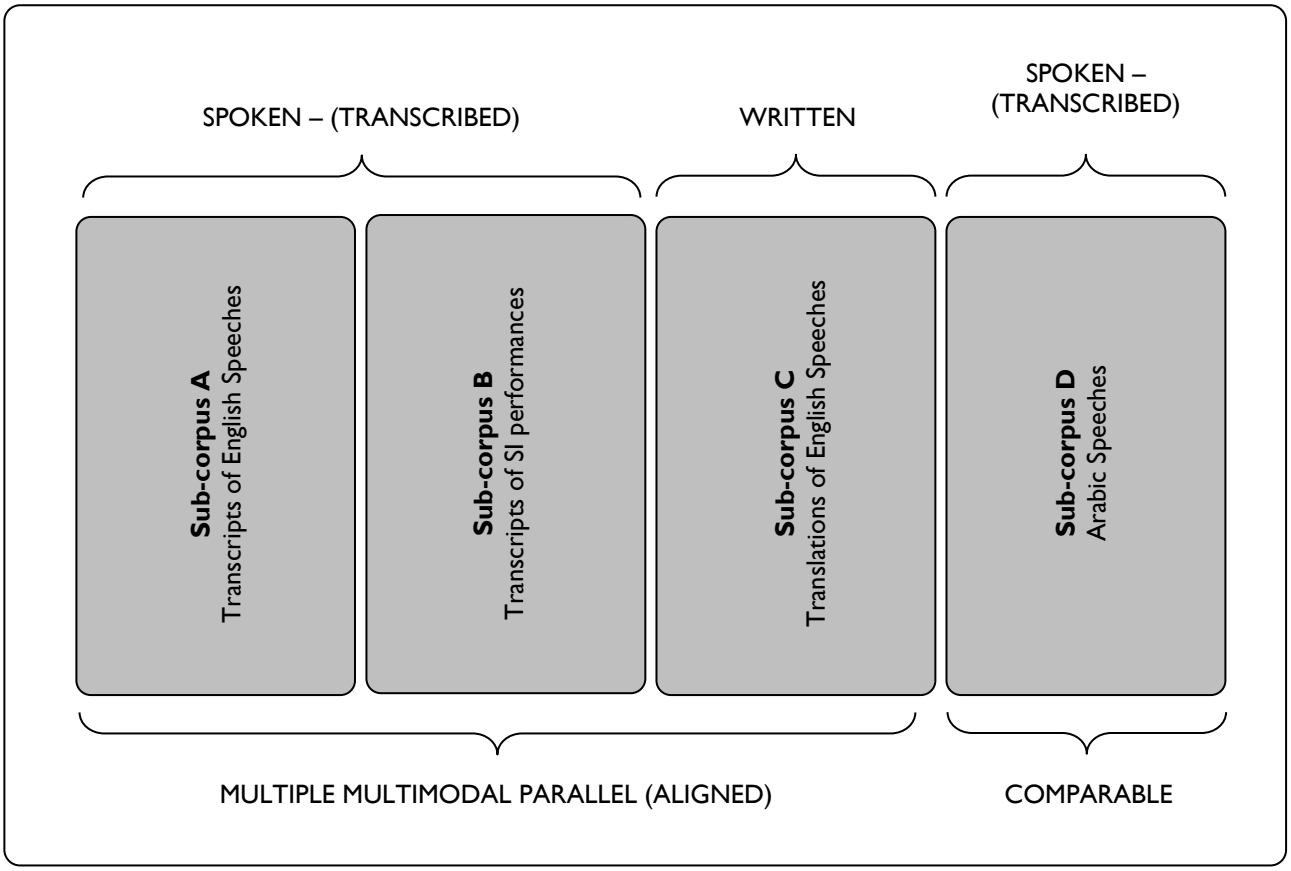
A corpus-based analysis is employed in this study as the primary methodology for examining contrastive aspects (Alonso-Almeida & Díez Abadie, 2025; Freitag & Rebechi, 2025), aiding in the validation or invalidation of linguistic intuitions and hypotheses (Baker, 1993, 2000; Ebeling & Ebeling, 2013; Johansson, 2007, 2010; Li, 2017; Saldanha, 2011; Shlesinger, 1998).

2.1 Corpus design

The monitoring of sentence patterns across interpreted, translated, and original speeches necessitates a modular corpus design (see Li, 2017). Accordingly, our corpus comprises four separate sub-corpora, as illustrated in Figure 1.

A. Sub-corpus A: consisting of a transcription of original English political speeches.
B. Sub-corpus B: consisting of professional English>Arabic SIs of the original English speeches in Sub-corpus A.
C. Sub-corpus C: consisting of Arabic translations of the original English speeches.
D. Sub-corpus D: consisting of transcribed original Arabic political speeches.

Figure 1: Corpus design



Source: Authors

Sub-corpora A, B, and C represent the multiple multimodal parallel component of our corpus. This component is multiple because several interpretations and translations into the same TL are available for a single source speech (Alonso Bacigalupe, 2010). This design departs from conventional parallel corpora, which typically follow a one-to-one design (Castagnoli, 2011; Li, 2017; Xia & Yarowsky, 2017) and offers the distinct advantage of enabling comparisons of renditions across interpreters, across translators, and between interpreters and translators (Castagnoli, 2011). This component of the corpus is equally multimodal, as it integrates both spoken (SI) and written (translation) performances. This design enables the identification of features specific to SI as a distinct mode (Bernardini et al., 2016; Kajzer-Wietrzny et al., 2022; Lefer & De Sutter, 2022). Additionally, this component is parallel, as its three sub-corpora are aligned, allowing for 'one-to-many' results (Castagnoli, 2011).

The fourth Sub-corpus (D) is a comparable reference component that comprises transcripts of original Arabic speeches. It serves as a reference point to account for patterns observed in the translating and interpreting performances (Defrancq & Plevoets, 2018). This section of the corpus claims comparability with the multiple multimodal parallel component in terms of domain and pragmatic function (political speeches) (Johansson, 2010), period covered (Laviosa, 1997), and sample size (Baker, 1995). Utilizing a combination of parallel and comparable sub-corpora, the overall corpus structure adheres to established recommendations in corpus-based studies of translation and interpreting (Baker, 1993, 1995, 2000; Johansson, 2010; Kajzer-Wietrzny et al., 2022; Kenning, 2010; Shlesinger, 1998). Similar designs have been employed in projects such as those by Bernardini and Zanettin (2004), Laviosa (2007), Munday (1998), and Freitag and Rebechi (2025).

2.2 Material selection and sampling

2.2.1 Sub-corpus A

The texts comprising Sub-corpus A consist of transcripts of seven authentic English political speeches, representative of those encountered in real-life conference interpreting settings (see Appendix I for detailed specifications). These speeches belong to the specific genre of political speeches. Since the language sampled is very specialized, fewer issues are expected to be encountered in text selection (Atkins et al., 1992). Thus, any English political speech delivered in a formal or conference-like setting, whether at the national and/or international level, and for which a professional Arabic SI is available is a candidate for inclusion in Sub-corpus A. All the speeches making up Sub-corpus A were prepared speeches delivered with a formal style typical in high-level international conferences. The presentation rates measured range from challenging and fast input rates of 145, 134, and 130 wpm on average in speeches V, IV and VI, respectively, to optimal and reasonably comfortable speech rates of 119, 115, 114, and 108 wpm on average in speeches I, VII, II, and III, respectively. All speeches delivered addressed topics and issues that are familiar to the professional interpreters.

2.2.2 Sub-corpus B

This section contains transcripts in Arabic, featuring multiple SIs of the English speeches from Sub-corpus A. Following an "opportunistic approach" to data collection (Love et al., 2017, p. 327), these performances were mined from various Arabic TV news channels that broadcasted these speeches live and subsequently made video recordings of these events available on their websites. Two performances were also sourced from the UN Web TV (see Appendix II for detailed specifications). We can assume with reasonable confidence that these interpreters had both the experience and expertise necessary for performing on such top-level conference settings, given that they are working either for the UN (Speeches II and IV) or for some of the most prominent Arabic TV news channels (all other speeches). All spoke with a native Arabic accent and appeared to be working under the psychological pressure characteristic of SI conditions due to such speech phenomena that marked their performance as pauses, hesitations, errors, false starts, ungrammatical and/or incomplete utterances (Chmiel et al., 2022).

2.2.3 Sub-corpus C

Sub-corpus C comprises written Arabic translations of the speeches from component A. Translations for speeches I, II, and III were sourced from various news and government agency websites. Translations for speeches IV to VII were unavailable and had to be commissioned, following recommendations in Johansson (2004). These commissioned translations were carried out by independent professional translators working at a well-known translation agency. The procurement order did not include specific instructions; translators were informed that the output might be used for research purposes but were not briefed on the project's topic (see Appendix III for details).

2.2.4 Sub-corpus D

Sub-corpus D comprises transcripts of 24 original Arabic political speeches delivered by high-level dignitaries. To satisfy the representativeness requirement (Biber, 1993; Halverson, 1998; Tognini-Bonelli, 2001), the selection criteria stipulate the following:

(1) Any Arabic political speech delivered by a native Arab official using Modern Standard Arabic (MSA) is a candidate for inclusion in the corpus.

(2) Additionally, the original Arabic speeches must not have been delivered earlier than 1970. This span of slightly over 50 years was adopted to provide broad temporal coverage, representativeness, and historical evidence of the consistency of VSO as the default structure in MSA, while still allowing for some overlap in period with the English speeches.

The overall size of this sub-corpus is approximately equivalent to that of component B (see Appendix IV for metadata of these speeches). At its current stage of development, the corpus comprises 112,821 words and 18h 27' 12" of video (see Table 1 for a summary).

Table 1: Multiple multimodal parallel and comparable corpus size

Materials and sample size				
Sub-corpora	**No. of texts**	**No. of words**	**No. of segments analyzed**	**Duration**
Sub-corpus A: Transcripts of English Speeches	7	16,260	NA	2h 18' 49"
Sub-corpus B: Transcripts of Arabic SI performances	21	42,332	3,113	8h 03' 09"
Sub-corpus C: Arabic translations of English speeches	8	16,555	1,208	NA
Sub-corpus D: Arabic original speeches	24	37,674	2,229	8h 05' 14"
Totals	**60**	**112,821**	**6,550**	**18h 27' 12"**

Source: Authors

2.3 Data coding

Manual verbatim orthographic transcriptions were generated for sub-corpora A, B, and D. The transcription approach followed a basic method deemed suitable for achieving the project's objectives. In addition to standard orthography and punctuation, speech phenomena such as hesitation, false starts, and pauses were faithfully included. These speech phenomena were transcribed without normalization. No XML markup was included in the current stage of corpus development. Additionally, no time alignment, synchronicity, or other temporal and prosodic features were introduced. The segmentation process was carried out manually, focusing on cohesive units at both syntactic and semantic levels (Glenn et al., 2010). Subsequently, an alignment phase followed, which was conducted manually due to the potential need for extensive editing if automatic alignment were relied upon. In addition to the remoteness of the language-pair, automatic alignment accuracy rates would have been affected by the blurry sentence boundaries, false starts, skipped and merged sentences, and other types of noise that usually characterize speech (Ma, 2006). While part-of-speech (POS) tagging is under consideration for future corpus developments, it is deemed unnecessary for the current investigation's objectives (Kenning, 2010). The corpus underwent several quality assurance checks, encompassing transcription orthography and tagging. The segments in sub-corpora B, C, and D were tagged for sentence structure, utilizing a binary categorization, 'VSO' for verb-initial and 'SVO' for subject-initial structures, respectively. Due to the oral nature of Sub-corpus B, certain segments included non-clause fragments, omissions, or incomplete sentences, which constitute noise in the data. These segments were classified as 'Other' and were excluded from the analysis.

2.4 Data analysis

Analysis was conducted on sub-corpora B, C, and D, with sub-corpus A excluded from the study. Thus, the investigation does not encompass a contrastive component between source and target languages. Instead, the study aims to explore the impact on Arabic sentence patterns across interpreted, translated, and originally Arabic texts, along lines advocated by Baker (1993, 1995) and Laviosa (1998). This allows for comparisons between translators' decisions in written texts and

interpreters' decisions in mediated communication, on the one hand, and with the reference sub-corpus containing original Arabic speeches in non-mediated communication, on the other.

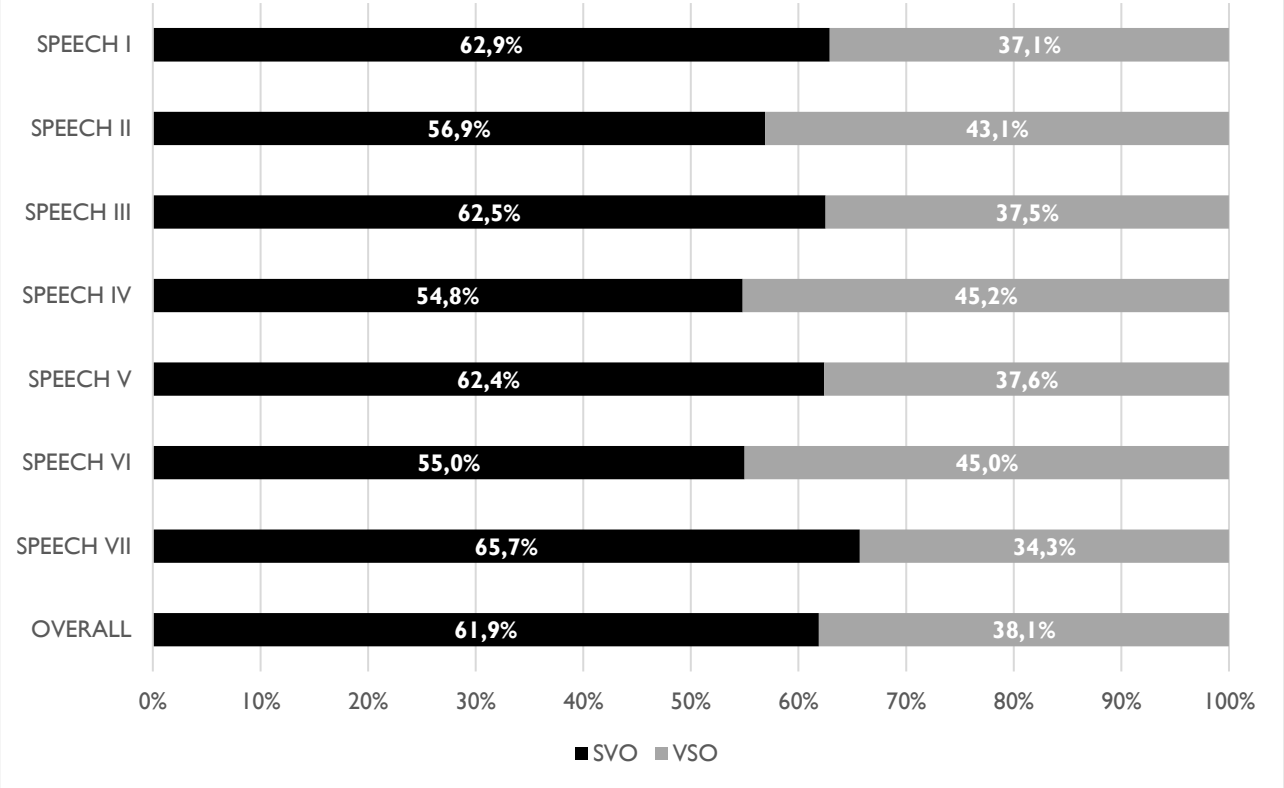
3. Results

In this section, we present the findings of the analysis of data concerning the SVO/VSO ratios in Sub-corpora B, C and D. We also provide the results regarding the relation between the type of mediation (SI as opposed to translation) and language modality (spoken *vs*. written) and choice of structure.

3.1 SVO/VSO ratio in Sub-corpus B

As Table 1 indicates, Sub-corpus B comprises 21 Arabic SI versions of the seven English speeches in Sub-corpus A. These have rendered a total of 3,113 segments that have been coded and analyzed for the type of sentence structure used by the 21 Arabic simultaneous interpreters, that is, either the default, unmarked VSO or derived and marked SVO structure.

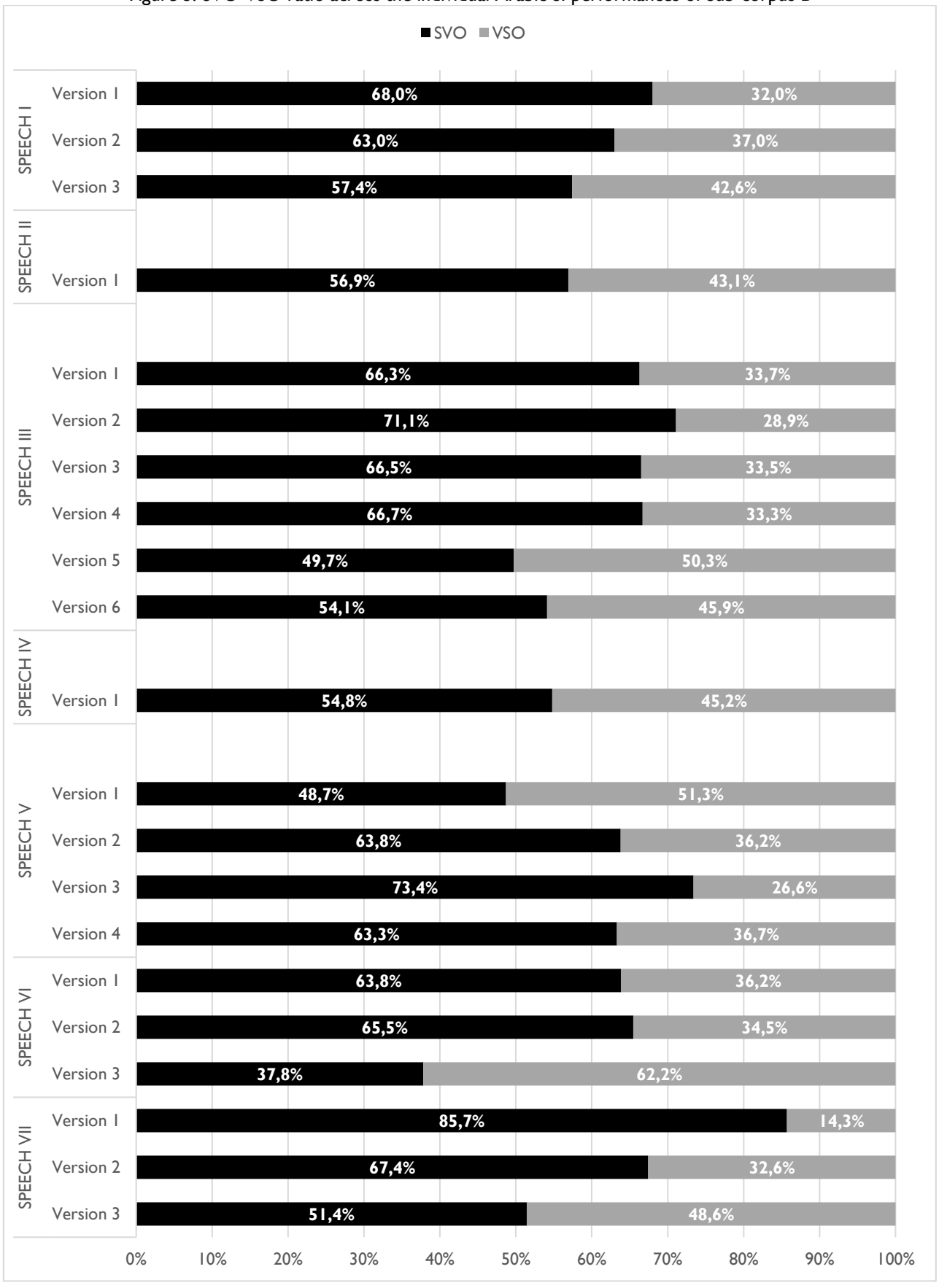
Figure 2: Summary of SVO/VSO ratio in Sub-corpus B



Source: Authors

As anticipated, an emerging trend is clearly observable in Figure 2, indicating that the SVO structure has been used more frequently by the Arabic simultaneous interpreters in most of the segments with an overall score of 61.9%. By contrast, the VSO structure was opted for in only 38.1% of the cases.

Figure 3: SVO-VSO ratio across the individual Arabic SI performances of Sub-corpus B



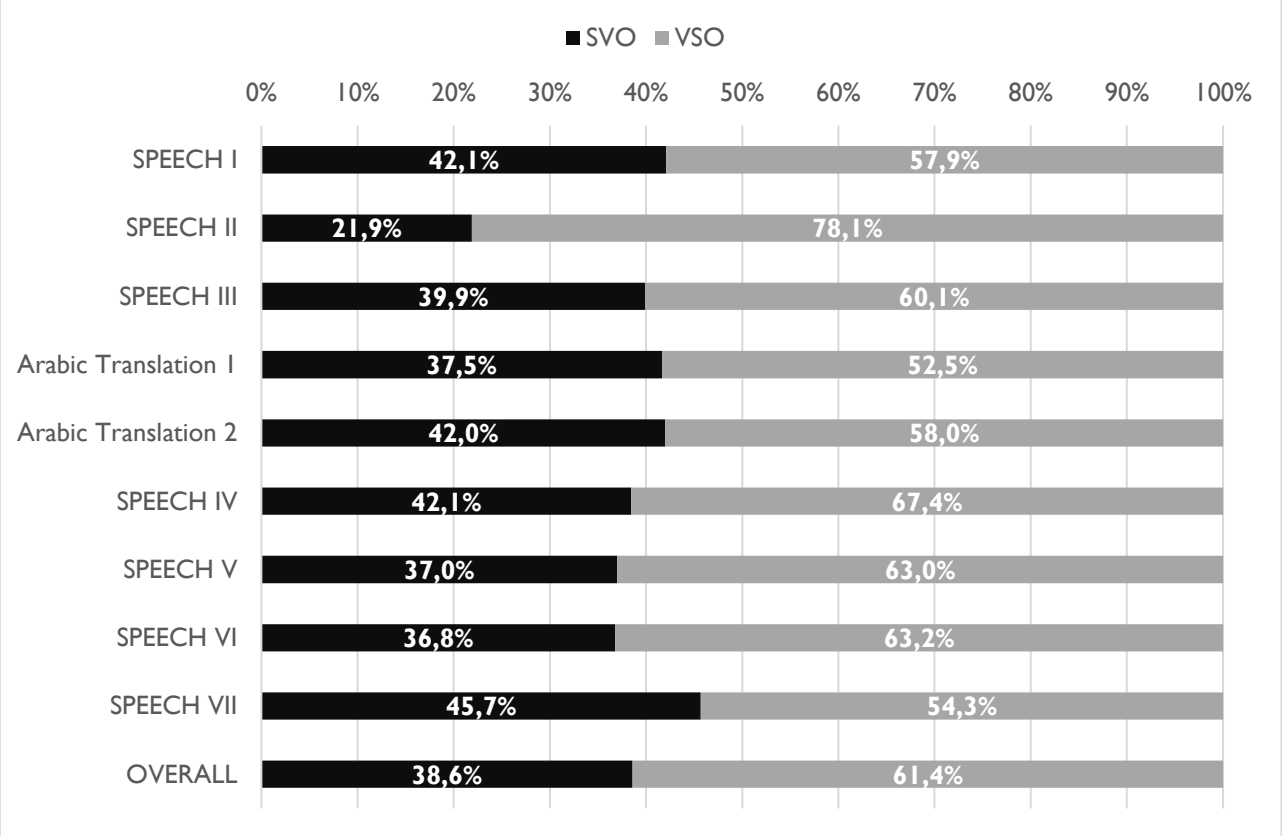
Source: Authors

Interestingly, almost comparable results have been obtained at the level of the individual Arabic SI performances making up Sub-corpus B with very few exceptions. Apart from the insignificant discrepancies observed in Versions 5 and 1 of Speeches III and V, respectively, and the obvious outlier shown in Version 3 of Speech VI, Figure 3 demonstrates that the SVO structure has been the structure of choice in most of the individual Arabic SI performances, ranging from a low score of 51.4% obtained in Version 3 of Speech VII to a remarkable 85.7% acquired in Version 1 of the same speech. A closer inspection of the outlier result observed in Version 3 of Speech VI revealed that the simultaneous interpreter employed a padding technique using the verb "لعرف" (*naʿrifu*, we know), which is analyzed as a VSO structure, at the beginning of the sentences in 12 instances out of the 51 VSO segments. Remarkably, no traces of "we know" whatsoever have been observed in the corresponding SL segments. It is expected that the result would have been reversed if that padding technique had not been used or had an SVO structure been used instead. What is certain, however, is that the use of this padding technique, an SVO structure or any other coping tactic is indicative of a certain type of difficulty faced by the interpreter and most probably brought about by structural asymmetry.

3.2 SVO/VSO ratio in Sub-corpus C

As indicated in the Materials and methods section, Sub-corpus C comprises written translations of the English conference speeches making up Sub-corpus A. In total, Sub-corpus C contains eight translations and 1,208 segments qualifying for analysis.

Figure 4: Summary of the SVO/VSO ratio in Sub-corpus C

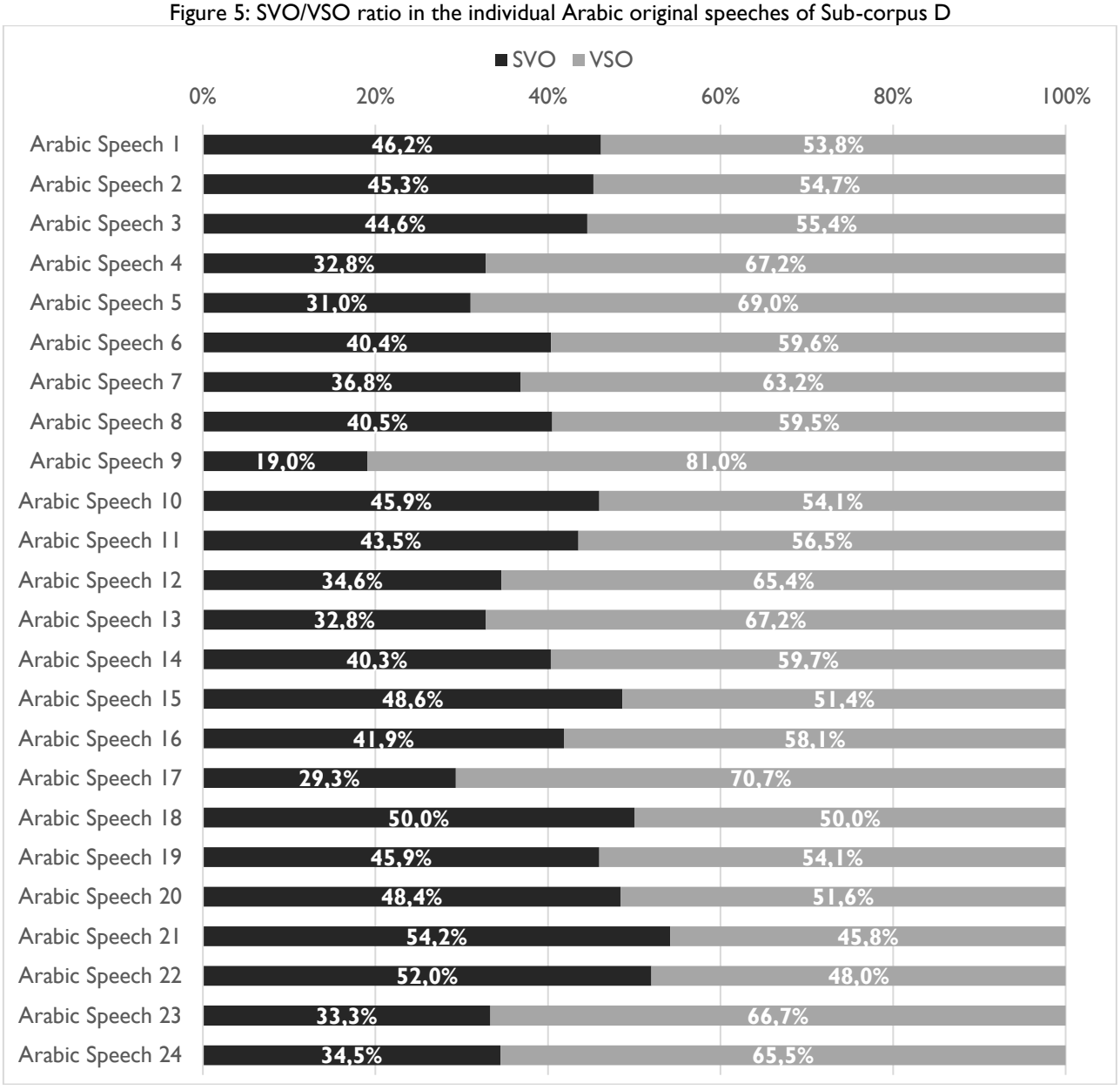


Source: Authors

As predicted, the trend emerging from the analysis of SVO/VSO ratio in Sub-corpus C goes in the reverse direction compared to the one obtained through the analysis of Arabic SI performances of Sub-corpus B. Figure 4 demonstrates that the VSO structure is the dominant one as it has consistently achieved higher scores than the SVO structure overall (at the sub-corpus and speech levels) and across the individual translations of Speech III with its two available translated versions. The results of Sub-corpus C analysis should be contrasted with the findings in Sub-corpus B and compared with our hypothesis stated in the introduction regarding the SVO/VSO ratio in interpreted speech versus translated texts. Simply stated, translators are more likely to opt for the basic and default V-initial structure in written discourse as they are not under the same pressure faced by simultaneous interpreters.

3.3 SVO/VSO ratio in Sub-corpus D

Figure 5: SVO/VSO ratio in the individual Arabic original speeches of Sub-corpus D

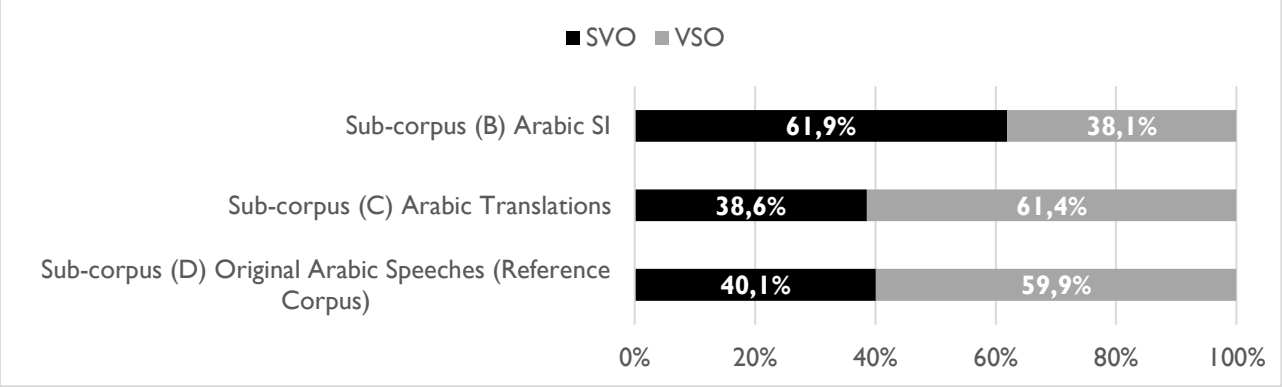


Source: Authors

Sub-corpus D, our reference corpus, includes a total of 24 original Arabic speeches and 2,229 segments. It is readily apparent from Figure 5 that there is a marked trend indicating that the VSO structure is predominant. This observation stands whether the results are approached at the sub-corpus level or at the level of individual speeches despite the apparent range and the obvious outliers of Speeches 21 and 22. Further analysis of the sentence patterns of these two speeches revealed that most of the sentences with an SVO structure or nominal clause contained a focus on the subject or a form of emphasis using the verb-like particle إن (*inna*), which means 'truly/surely' and which is used in a nominal clause. In both cases, the use of an SV(O) structure or nominal clause is expected. The findings of the analysis of Sub-corpus D are consistent with those in Sub-corpus C as well as with our hypothesis. As expected, however, these findings stand in stark contrast to the SVO-VSO ratio in Sub-corpus B.

A summary of the results, as demonstrated in Figure 6, provides a comparison of the findings obtained regarding the SVO-VSO ratio across all three sub-corpora. The figure indicates that the SVO-VSO ratios in Sub-corpus B on the one hand and Sub-corpora C and D on the other hand sit at opposite ends. In other words, the SVO structure is the dominant one in Sub-corpus B whereas the VSO structure dominates in Sub-corpora C and D. This marked trend is also maintained for the most part at the level of speeches and individual SI performances as shown in the breakdown of the results in the previous sub-sections.

Figure 6: Summary of SVO/VSO ratio in Sub-corpora B, C and D

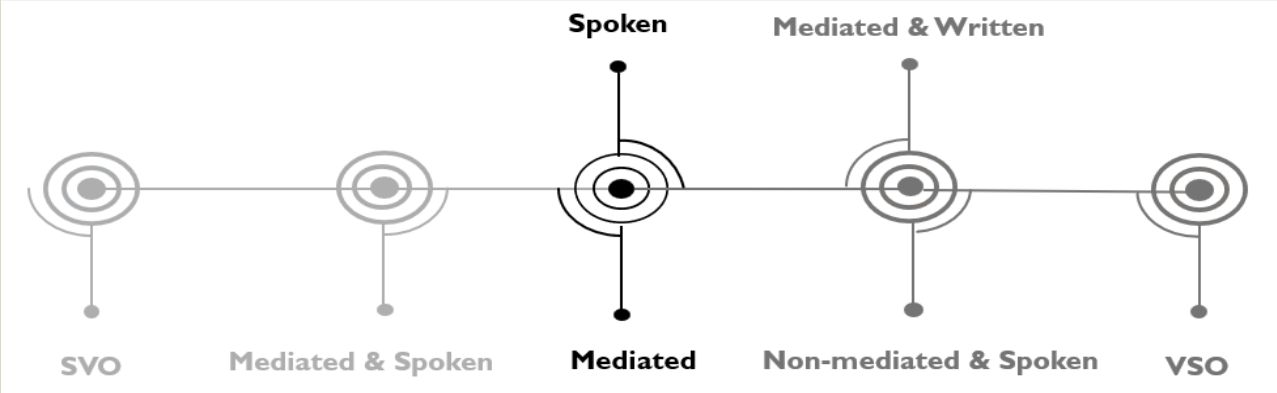


Source: Authors

3.4 Type of mediation and language modality and choice of structure

Our findings indicate that the choice of structure was not based on a single mediation type (SI *vs.* translation) or language modality (spoken *vs.* written), but a combination thereof. Thus, if our sub-corpora were placed on a continuum with the SVO structure at one extreme and the VSO structure at the other (Figure 7), we would observe that the more mediated and spoken the discourse, the closer it will be to the SVO structure. Conversely, the more mediated and written or non-mediated and spoken the discourse, the closer it will be to the VSO structure.

Figure 7: Choice of structure across language modalities



Source: Authors

4. Discussion

The overall results obtained from the analysis of the three sub-corpora provide clear answers to our research questions regarding the comparison of the SVO-VSO ratios across simultaneously interpreted, translated and original speeches. Whereas the dominant structure in Sub-corpus B has been the S-initial structure, the V-initial structure has been the dominant one both in Sub-corpus C and Sub-corpus D.

The choice of the S-initial structure in Sub-corpus B points to a coping tactic by the Arabic simultaneous interpreters to opt for the marked and derived S-initial structure that is easier to process due to its similarity to the SL English structures. Opting for the S-initial, though marked and derived, structure helps the interpreters closely follow the SL discourse structure without the need to lag far behind the SL input or overload their memory, which would otherwise occur if a V-initial structure was used through restructuring. It is intriguing that this choice of the S-initial structure by the Arabic simultaneous interpreters comes despite the risk of shift in emphasis involved in opting for the marked S-initial structure more often.

These findings are consistent with our hypothesis stated in Section 1 regarding the SVO/VSO ratio in interpreted speech versus translated and original texts. Moreover, the SVO-VSO ratios in Sub-corpus C and Sub-corpus B are in complete agreement with the well-established evidence that the VSO structure is the basic, default or dominant word order in original Arabic discourse (Dahlgren, 2009; Fassi-Fehri, 1993; Hoyt, 2009; Mohammad, 2000).

The result regarding the relation between the type of mediation and language modality on the one hand and the choice of structure on the other provides further evidence, pointing to the impact of these two variables on the choice of structure and at the same time relating the structural choices by the interpreters to the SI process and syntactic asymmetry.

Our current findings provide concrete evidence based on which we can comfortably relate the interpreters' choices to the SI process and also to syntactic asymmetry. Being under the constant psychological pressure characteristic of SI and faced with syntactic dissimilarity, the simultaneous interpreters are highly likely to opt for the structure that can be easily processed and the one that is less likely to lead to prolonged time lag, cognitive overload, and potentially associated risks. Arabic, with its flexible word order, offers the S-initial structure that is similar to the SL English word order

and allows the Arabic simultaneous interpreters to follow the SL structure closely. In so doing, the simultaneous interpreters stand a greater opportunity to avoid the potential risks associated with restructuring using the V-initial structure, including prolonged time lag, storage of excessive amounts of information, omissions or complete communication failure at times. Accordingly, the simultaneous interpreters' choice of the S-initial structure can be viewed as a coping tactic to overcome structural disparity and potential cognitive overload, information loss or complete failure.

It can thus be concluded that the Arabic simultaneous interpreters were in effect adopting a "form-based" (Isham, 1994, p. 207-208) or structure-based approach as opposed to meaning-based processing. This tendency is evidenced by the use of the S-initial structure that is similar or identical to the SL English structure in a bid to avoid restructuring and waiting for the verb and the potential risks associated with these two strategies. By closely following the SL structure, the Arabic simultaneous interpreters can easily follow the original speaker with reduced time lag and no need for storing large amounts of input chunks, thereby avoiding potential cognitive overload and information loss. While external influences cannot be entirely ruled out, the triangulated corpus evidence in this study isolates syntactic asymmetry as the most plausible factor driving this structural shift. In comparable original Arabic speeches and in written translations of the same English conference speeches, VSO dominates, reflecting the expected norm in MSA and the register typically used in international conferences. By contrast, in SI, the interpreters' consistent preference for SVO points to the real-time need to align with the English SVO input and mitigate processing load. Adopting SVO over VSO produces a marked change in information structure and emphasis, which would normally be avoided in the formal register of conference Arabic. Taken together, these considerations suggest that the interpreters' tendency to opt for SVO is best explained by structural asymmetry and the processing demands of SI.

Our findings corroborate previous evidence reported on syntactic segmentation (Dayter, 2020; Goldman-Eisler, 1972) and the lack of need or support for re-ordering when the language pairs have symmetrical syntactic constructions (Al-Rubai'i, 2004; Alonso Bacigalupe, 2010; Chen et al., 2015; El-Zawawy, 2021; Russo, 1997; Setton & Motta, 2007). They also confirm our earlier findings on the same topic (Al Zahran, 2021; Al Zahran & Jamoussi, 2022).

More importantly, our findings are hardly distinguishable from a significant body of literature, which provided evidence substantiating the language-pair specificity hypothesis as reported in Sub-section 1.2. The ease of processing brought about by Arabic flexible word order and its S-initial structure which is similar or identical to the SL English structures is not available to other interpreters working with other language pairs that have asymmetrical structures or languages with rigid structures such as German (Bevilacqua, 2009). Nevertheless, although our results support the language-pair specificity hypothesis, our corpus-based analysis provides evidence that attests to the plausibility of replicating this syntactic alignment in other language pairs that have structures leading to similar SI challenges or that have syntactic asymmetry combined with flexible word order.

Our findings demonstrate that the professional interpreters were prioritizing completeness over discourse and emphasis shifts to cope with the pressure of the SI process and syntactic asymmetry. They can thus provide insights that could readily be applied in interpreter pedagogy and professional development courses. By exposing student interpreters to the challenges of structural asymmetry and the potential problems associated with it, interpreter trainers can provide trainees

with the opportunity to deploy this coping tactic, especially when syntactic disparity coincides with other extreme factors such as information density, syntactic complexity, high SL input rate, and so on. Targeted exercises could involve guided practice on authentic English speeches to produce SVO renderings under both normal and high-load conditions, reinforcing the benefits of SVO alignment in reducing lag and memory strain while preserving accuracy, completeness, and fluency. Other drills may include comparative re-rendering tasks of identical segments in both the marked SVO and default VSO forms to assess processing efficiency and communicative effect. Integrating such activities into interpreter-training curricula can supplement activities that are meant to strengthen trainees' ability to manage structural asymmetry challenges in real time. For further pedagogical applications of syntactic asymmetry in interpreter training, see (Al Zahran, 2021).

At the same time, the findings call for a debate among Arabic linguists, interpreting scholars and interpreter trainers, as well as professional interpreters on the tolerance of the S-initial structure in standard Arabic during SI when the expected sentence pattern is the default and dominant V-initial structure.

Finally, our findings and conclusions should be approached while bearing in mind the following limitation. Although our corpus is substantial in the context of interpreting studies, it can still be enhanced in terms of size and broader language-pair coverage. Further research on larger corpora involving more language pairs and directions would be needed to corroborate our conclusions with more conclusive evidence.

5. Conclusion

This article has examined one aspect of language-pair specificity, namely the SVO-VSO syntactic asymmetry in English>Arabic SI. A corpus-based analysis has been conducted on three types of datasets, including simultaneously interpreted, translated and original Arabic discourse. The analysis has revealed that while the Arabic simultaneous interpreters resorted to the S-initial structure more frequently as a tactic to cope with SI pressure and syntactic asymmetry, the V-initial structure was the dominant one in translated and original Arabic spoken discourse. The analysis has also revealed that the choice of structure was affected by the interplay between the type of mediation and language modality, that is, mediated and spoken (SI) as opposed to mediated and written (translation) on the one hand, and non-mediated and spoken on the other (original Arabic spoken discourse). The effects of these variables help further relate the current findings to the SI process and impact of syntactic asymmetry. We have thus concluded that the Arabic simultaneous interpreters were applying form-based as opposed to meaning-based processing, as evidenced by the use of the Arabic S-initial structure, which is similar or identical to SL structures. We have also concluded that our study has provided evidence in favor of the language-pair specificity hypothesis.

References

Abdul-Raof, H. (1998). *Subject, Theme and Agent in Modern Standard Arabic*. Routledge. https://doi.org/10.4324/9780203036730

Al Zahran, A. (2021). Structural Challenges in English>Arabic Simultaneous Interpreting. *Translation & Interpreting, 13*(1), 51–70. https://doi.org/10.12807/ti.113201.2021.a04

Al Zahran, A. & Jamoussi, R. (2022). A Corpus-based Investigation of VSO-SVO Usage in Simultaneous Interpreting. *Hikma: Translation Studies Journal*, *21*(2), 231–255. https://doi.org/10.21071/hikma.v21i2.14281

Al-Jarf, R. (2022). Student-interpreters' Foreign Proper Noun Pronunciation Errors in English-Arabic and Arabic-English Media Discourse Interpreting. *International Journal of Translation and Interpretation Studies*, *2*(1), 80–90. https://doi.org/10.32996/ijtis.2022.2.1.11

Alonso-Almeida, F. J., & Díez Abadie, G. (2025). Evidential English Adverbials and their French Equivalents in a Specialised Parallel Corpus. *Cadernos de Tradução*, *45*, 1–23. https://doi.org/10.5007/2175-7968.2025.e99785

Alonso Bacigalupe, L. (2010). Information Processing During Simultaneous Interpretation: A Three-tier Approach. *Perspectives*, *18*(1), 39–58. https://doi.org/10.1080/09076760903464278

Al-Rubai'i, A. (2004). The Effect of Word Order Differences on English-into-Arabic Simultaneous Interpreters' Performance. *Babel*, *50*(3), 246–266. https://doi.org/10.1075/babel.50.3.04alr

Aoun, J., Benmamoun, E., & Sportiche, D. (1994). Agreement, Word-order, and Conjunction in some Varieties of Arabic. *Linguistic Inquiry, 25*(2), 195–220.

Atkins, S., Clear, J., & Oster, N. (1992). Corpus Design Criteria. *Literary and Linguistic Computing*, *7*(1), 1–16. https://doi.org/10.1093/llc/7.1.1

Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). John Banjamins.

Baker, M. (1995). Corpora in Translation Studies. *Target*, *7*(2), 223–243. https://doi.org/10.1075/target.7.2.03bak

Baker, M. (2000). Towards a Methodology for Investigating the Style of a Literary Translator. *Target*, *12*(2), 241–266. https://doi.org/10.1075/target.12.2.04bak

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

Bernardini, S., & Zanettin, F. (2004). When is a Universal not a Universal? In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do they Exist?*. (pp. 51–62). John Benjamins. https://doi.org/10.1075/btl.48.05ber

Bevilacqua, L. (2009). The Position of the Verb in Germanic Languages and Simultaneous Interpretation. *The Interpreters' Newsletter*, *14*, 1–31.

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, *8*(4), 243–257.

Castagnoli, S. (2011). Exploring Variation and Regularities in Translation with Multiple Translation Corpora. *Rassegna Italiana di Linguistica Applicata*, *43*(1), 311–332.

Chang, C., & Schallert, D. L. (2007). The Impact of Directionality on Chinese/English Simultaneous Interpreting. *Interpreting*, *9*(2), 137–176. https://doi.org/10.1075/intp.9.2.02cha

Chen, Y., Song, Z., & Wu, C. (2015). Syntactic Linearity as a Strategy in Simultaneous Interpreting: A Case Study on English-Chinese Interpretation. *T&I Review*, *5*, 29–69.

Chmiel, A., Koržinek, D., Kajzer-Wietrzny, M., Janikowski, P., Jakubowski, D., & Polakowska, D. (2022). Fluency Parameters in the Polish Interpreting Corpus (PINC). In M. Kajzer-Wietrzny, A. Ferraresi, I. Ivaska & S. Bernardini (Eds.), *Mediated Discourse at the European Parliament: Empirical Investigations* (pp. 63–91). Language Science Press.

Chmiel, A., Kajzer-Wietrzny, M., Koržinek, D., Jakubowski, D., & Janikowski, P. (2024). Syntax, Stress and Cognitive Load, or on Syntactic Processing in Simultaneous Interpreting. *Translation, Cognition & Behavior*, *7*(1), 22–47. https://doi.org/10.1075/tcb.00091.chm

Collard, C., Przybyl, H., & Defrancq, B. (2018). Interpreting into an SOV Language: Memory and the Position of the Verb: A Corpus-based Comparative Study of Interpreted and Non-mediated Speech. *Meta*, *63*(3), 695–716. https://doi.org/10.7202/1060169ar

Dahlgren, S.-O. (2009). Word Order. In K. Versteegh (Ed.), *Encyclopedia of Arabic language and Linguistics* (vol. IV, pp. 725–736). Koninklijke Brill NV.

Dawrant, A. C. (1996). *Word Order in Chinese to English Simultaneous Interpretation: An Initial Exploration*. [Unpublished Master's Thesis]. Fu Jen Catholic University.

Dayter, D. (2020). Strategies in a Corpus of Simultaneous Interpreting: Effects of Directionality, Phraseological Richness, and Position in Speech Event. *Meta*, *65*(3), 594–617. https://doi.org/10.7202/1077405ar

Defrancq, B., & Plevoets, K. (2018). Over-uh-load, Filled Pauses in Compounds as a Signal of Cognitive Load. In M. Russo, C. Bendazzoli & B. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies* (pp. 43–64). Springer. https://doi.org/10.1007/978-981-10-6199-8_3

Ebeling, S. O., & Ebeling, J. (2013). From Babylon to Bergen: On the Usefulness of Aligned Texts. *Bergen Language and Linguistics Studies*, *3*(1), 23–42. https://doi.org/10.15845/bells.v3i1.359

El-Zawawy, A. M. (2021). Simultaneous Interpretation of Complex Structures from English into Arabic. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, *35*(1), 30–64. https://doi.org/10.1075/resla.19053.zaw

Fassi-Fehri, A. (1993). *Issues in the Structure of Arabic Clauses and Words*. Springer. https://doi.org/10.1007/978-94-017-1986-5

Freitag, P. H., & Rebechi, R. (2025). Analysing Collocations in Articles Published in Coworking Blogs in Portuguese: A Comparative Study on Authentic and Translated Texts. *Cadernos de Tradução*, *45*, 1–25. https://doi.org/10.5007/2175-7968.2025.e102039

Gile, D. (1991). The Processing Capacity Issue in Conference Interpretation. *Babel*, *37*(1), 15–27. https://doi.org/10.1075/babel.37.1.04gil

Gile, D. (1992). Predictable Sentence Endings in Japanese and Conference Interpretation. *The Interpreters' Newsletter*, *Special Issue*(1), 12–23.

Gile, D. (1999). Testing the Effort Models' Tightrope Hypothesis in Simultaneous Interpreting: A Contribution. *Hermes*, *12*(23), 153–172. https://doi.org/10.7146/hjlcb.v12i23.25553

Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training* (rev. ed.). John Benjamins.

Gile, D. (2011). Errors, Omissions and Infelicities in Broadcast Interpreting: Preliminary Findings from a Case Study. In C. Alvstad, A. Hild & E. Tiselius (Eds.), *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies* (pp. 201–218). John Benjamins.

Glenn, M. L., Strassel, S. M., Lee, H., Maeda, K., Zakhary, R., & Li, X. (2010). Transcription Methods for Consistency, Volume and Efficiency. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner & D. Tapias (Eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (pp. 2915–2920). European Language Resources Association.

Goldman-Eisler, F. (1972). Segmentation of Input in Simultaneous Translation. *Journal of Psycholinguistic Research*, *1*(2), 127–140. https://doi.org/10.1007/bf01068102

Guo, L. (2011). *An Analysis of the Word Order Pattern in the SI Target Language and its Underlying Reasons in the Language Combination of English and Chinese.* [Unpublished Doctoral Dissertation]. Shanghai International Studies University.

Halverson, S. (1998). Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical/Descriptive Categories and a Conception of the Object of Study. *Meta*, *43*(4), 494–514. https://doi.org/10.7202/003000ar

He, H., Boyd-Graber, J., & Daumé III, H. (2016). Interpretese vs. Translationese: The Uniqueness of Human Strategies in Simultaneous Interpretation. In K. Knight, A. Nenkova & O. Rambow (Eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 971–976). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-1111

Hoyt, F. M. (2009). Verbal Clause. In K. Versteegh (Ed.), *Encyclopedia of Arabic Language and Linguistics* (pp. 653–659). Brill.

Isham, W. P. (1994). Memory for Sentence Form after Simultaneous Interpretation: Evidence both for and against Deverbalization. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the Gap: Empirical Research in Simultaneous Interpretation* (pp. 191–211). John Benjamins.

Johansson, S. (2004). Why Change the Subject? *Target*, *16*(1), 29–52. https://doi.org/10.1075/target.16.1.03joh

Johansson, S. (2007). Seeing through Multilingual Corpora. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years on* (pp. 51–71). Rodopi.

Johansson, S. (2010). Multilingual Corpora: Possibilities and Limitations. In F. Čermák, P. Corness & A. Klégr (Eds.), *InterCorp: Exploring a Multilingual Corpus* (pp. 53–68). Lidové Noviny.

Kade, O., & Cartellieri, C. (1971). Some Methodological Aspects of Simultaneous Interpreting. *Babel*, *17*(2), 12–16. https://doi.org/10.1075/babel.17.2.04car

Kajzer-Wietrzny, M., Ferraresi, A., Ivaska, I., & Bernardini, S. (2022). Using European Parliament Data in Translation and Interpreting Research: An Introduction. In M. Kajzer-Wietrzny, A. Ferraresi, I. Ivaska & S. Bernardini (Eds.), *Mediated Discourse at the European Parliament: Empirical Investigations* (pp. iii–xi). Language Science Press. https://doi.org/10.5281/zenodo.6977036

Kenning, M.-M. (2010). What are Parallel and Comparable Corpora and How Can We Use Them? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 487–500). Routledge.

Kirchhoff, H. (2002). Simultaneous Interpreting: Interdependence of Variables in the Interpreting Process, Interpreting Models and Interpreting Strategies. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies Reader* (pp. 111–119). Routledge.

Laviosa, S. (1997). How Comparable Can "Comparable Corpora" Be? *Target*, *9*(2), 287–319. https://doi.org/10.1075/target.9.2.05lav

Laviosa, S. (1998). The Corpus-based Approach: A New Paradigm in Translation Studies. *Meta*, *43*(4), 474–479. https://doi.org/10.7202/003424ar

Laviosa, S. (2007). Studying Anglicisms with Comparable and Parallel Corpora. *Belgian Journal of Linguistics*, *21*(1), 123–136. https://doi.org/10.1075/bjl.21.09lav

Lefer, M.-A., & De Sutter, G. (2022). Using the Gravitational Pull Hypothesis to Explain Patterns in Interpreting and Translation: The Case of Concatenated Nouns in Mediated European Parliament Discourse. In M. Kajzer-Wietrzny, A. Ferraresi, I. Ivaska & S. Bernardini (Eds.), *Mediated Discourse at the European Parliament: Empirical Investigations* (pp. 133–159). Language Science Press. https://doi.org/10.5281/zenodo.6977046

Li, D. (2017). Translator Style: A Corpus-assisted Approach. In M. Ji, L. Hareide, D. Li & M. Oakes (Eds.), *Corpus Methodologies Explained: An Empirical Approach to Translation Studies*. Routledge.

Liontou, K. (2011). Anticipation in German to Greek Simultaneous Interpreting. *Gramma*, *19*, 37–56. https://doi.org/10.26262/gramma.v19i0.6325

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations. *International Journal of Corpus Linguistics*, *22*(3), 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

Ma, X. (2006). Champollion: A Robust Parallel Text Sentence Aligner. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk & D. Tapias (Eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)* (489–492). European Language Resources Association.

Meuleman, C., & Van Besien, F. (2009). Coping with Extreme Speech Conditions in Simultaneous Interpreting. *Interpreting*, *11*(1), 20–34. https://doi.org/10.1075/intp.11.1.03meu

Mohammad, M. A. (2000). *Word Order, Agreement and Pronominalization in Standard and Palestinian Arabic*. John Benjamins. https://doi.org/10.1075/cilt.181

Moser, B. (1978). Simultaneous Interpretation: A Hypothetical Model and its Practical Application. In D. Gerver & H. W. Sinaiko (Eds.), *Language Interpretation and Communication* (pp. 353–368). Plenum Press.

Munday, J. (1998). A Computer-assisted Approach to the Analysis of Translation Shifts. *Meta*, *43*(4), 542–556. https://doi.org/10.7202/003680ar

Riccardi, A. (1996). Language-specific Strategies in Simultaneous Interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching Translation and Interpreting 3: New Horizons* (pp. 213–222). John Benjamins.

Rodríguez-Inés, P., & Gallego-Hernández, D. (2016). Corpus Use and Learning to Translate, Almost 20 Years on. *Cadernos de Tradução*, *36*(1), 9–13. https://doi.org/10.5007/2175-7968.2016v36nesp1p9

Russo, M. (1997). Morphosyntactical Assymmetries between Spanish and Italian and their Effect during Simultaneous Interpreting. In K. Klaudy & J. Kohn (Eds.), *Transferre necesse est: Proceedings of the 2nd International Conference on Current Trends in Studies of Translation and Interpreting 5-7 September, 1996, Budapest, Hungary* (pp. 268–272). Scholastica.

Saldanha, G. (2011). Translator Style: Methodological Considerations. *The Translator*, *17*(1), 25–50. https://doi.org/10.1080/13556509.2011.10799478

Seeber, K. G., & Kerzel, D. (2011). Cognitive Load in Simultaneous Interpreting: Model Meets Data. *International Journal of Bilingualism*, *16*(2), 228–242. https://doi.org/10.1177/1367006911402982

Seleskovitch, D., & Lederer, M. (1995). *A Systematic Approach to Teaching Interpretation*. (J. Harmer, Trans.). Registry of Interpreters for the Deaf.

Setton, R. (1993). Is Non-Intra-IE Interpretation Different? *Meta*, *38*(2), 238–256. https://doi.org/10.7202/004115ar

Setton, R. (1999). *Simultaneous Interpretation: A Cognitive-pragmatic Analysis*. John Benjamins. https://doi.org/10.1075/btl.28

Setton, R. (2005). So what is so interesting about Simultaneous Interpreting? *SKASE: Journal of Translation and Interpretation*, *1*(1), 70–84.

Setton, R., & Motta, M. (2007). Syntacrobatics. *Interpreting*, *9*(2), 199–230. https://doi.org/10.1075/intp.9.2.04set

Shamy, M., & de Pedro Ricoy, R. (2017). Retrospective Protocols: Tapping into the Minds of Interpreting Trainees. *Translation & Interpreting*, *9*(1). https://doi.org/10.12807/ti.109201.2017.a05

Shlesinger, M. (1998). Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies. *Meta, 43*(4), 486–493. https://doi.org/10.7202/004136ar

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins. https://doi.org/10.1075/scl.6

Van Besien, F. (1999). Anticipation in Simultaneous Interpretation. *Meta*, *44*(2), 250–259. https://doi.org/10.7202/004532ar

Wang, B., & Gu, Y. (2016). An Evidence-based Exploration into the Effect of Language-pair Specificity in English-Chinese Simultaneous Interpreting. *Asia Pacific Translation and Intercultural Studies*, *3*(2), 146–160. https://doi.org/10.1080/23306343.2016.1182238

Wang, B., & Zou, B. (2018). Exploring Language Specificity as a Variable in Chinese-English Interpreting: A Corpus-based Investigation. In M. Russo, C. Bendazzoli & B. Defrancq (Eds.), *Making way in Corpus-based Interpreting Studies* (pp. 65–82). Springer. https://doi.org/10.1007/978-981-10-6199-8_4

Wilss, W. (1978). Syntactic Anticipation in German-English Simultaneous Interpreting. In Gerver David & H. W. Sinaiko (Eds.), *Language Interpretation and Communication* (pp. 343–352). Plenum Press.

Xia, P., & Yarowsky, D. (2017). Deriving Consensus for Multi-Parallel Corpora: An English Bible Study. In G. Kondrak & T. Watanabe (Eds.), *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (pp. 448–453). Asian Federation of Natural Language Processing.

Zou, B., & Wang, B. (2023). Non-fluency and Language-pair Specificity in Chinese-English Consecutive Interpreting: A Corpus-driven Study. *Corpus-Based Translation Studies*, *11*(2), 30–49. https://doi.org/10.32714/ricl.11.02.03

Appendices

Appendix I. English conference speeches making up Sub-corpus A

Speech	Date delivered	Speaker	Occasion/Venue
Speech I	June 4, 2009	Former US President Barak Obama	Cairo University, Egypt
Speech II	January 31, 2014	UN Diplomat Lakhdar Brahimi	Written statement and question-and-answer session at the UN Geneva Office
Speech III	May 21, 2017	Former US President Donald Trump	Arab Islamic American Summit in Riyadh, Saudi Arabia
Speech IV	September 24, 2021	UN Secretary General António Guterres	High-Level Dialogue on Energy
Speech V	November 1, 2021	US President Joseph Robinette Biden Jr.	COP26, 26th Climate Summit, Glasgow, Scotland
Speech VI	November 1, 2021	Former UK Prime Minister Boris Johnson	COP26, 26th Climate Summit, Glasgow, Scotland
Speech VII	November 1, 2021	Prince Charles (now King) of the United Kingdom	COP26, 26th Climate Summit, Glasgow, Scotland

Appendix II. Sub-corpus B: Arabic SI versions of English speeches in Sub-corpus A

Speech #	SI Version	Broadcasting channel
Speech I	SI V1	Al Jazeera
	SI V2	Al Arabia
	SI V3	Egyptian Channel One
Speech II	SI V1	UN Web TV
Speech III	SI V1	Al Jazeera Mubasher
	SI V2	Al Saudia
	SI V3	Al Arabia
	SI V4	Al Hadath
	SI V5	Al Ghad
	SI V6	France 24 Arabic
Speech IV	SI V1	UN Web TV
Speech V	SI V1	Al Ghad TV
	SI V2	Al Hurra TV Channel
	SI V3	France Arabic 24
	SI V4	Sky News Arabia
Speech VI	SI V1	Asharq News
	SI V2	Extra News
	SI V3	Sky News Arabia
Speech VII	SI V1	Al Hayah TV
	SI V2	ON TV
	SI V3	Sky News Arabia
Total SI Versions	21	

Appendix III. Sub-corpus C: Arabic written translations of English speeches in Sub-corpus A

Speech #	Translation Version	Translation Source
Speech I	AR Trans. V1	Al Jazeera
Speech II	AR Trans. V1	UN Web TV
Speech III	AR Trans. V1	Al Jazeera Mubasher
	AR Trans. V2	Al Saudia
Speech IV	AR Trans. V1	UN Web TV
Speech V	AR Trans. V1	Al Ghad TV
Speech VI	AR Trans. V1	Asharq News
Speech VII	AR Trans. V1	Al Hayah TV
Total Arabic Translation Versions	**8**	

Appendix IV. Sub-corpus D: original Arabic speeches making up a comparative reference corpus

Speech	Date delivered	Speaker	Title and Occasion/Venue
Arabic Speech 1	Oct. 2, 2019	Princess Haya Bint Al Hussein	Future of Humanitarian Aid, World Government Summit, Dubai
Arabic Speech 2	Jul. 21, 2017	Tamim bin Hamad Al Thani Emir of Qatar	Status Que and Future Directions for the State of Qatar During the Gulf Crisis
Arabic Speech 3	Nov. 26, 1975	Qaboos bin Said, Sultan of Oman	A Speech to the People
Arabic Speech 4	Oct. 25, 2011	Sheikh Sabah Al-Ahmad, Emir of Kuwait	Speech of His Highness Emir Sheikh Sabah Al-Ahmad, Emir of Kuwait, before the opening session of the National Assembly at the opening of the fourth regular session of the thirteenth legislative term of the National Assembly
Arabic Speech 5	Nov. 1, 2022	Abdelmadjid Tebboune, President of Algeria	President Tebboune's speech at the opening of the 31st Arab Summit
Arabic Speech 6	Jul. 13, 2022	Sheikh Mohammed bin Zayed Al Nahyan, President of the United Arab Emirates	Sheikh Mohammed bin Zayed's speech to the people and residents of the UAE
Arabic Speech 7	Mar. 31, 2019	Salman bin Abdulaziz Al Saud, King of Saudi Arabia	King Salman's speech at the 30th Arab Summit - Tunisia
Arabic Speech 8	Jan. 24, 2018	Interior Minister of Egypt, Major General Magdy Abdel Ghaffar	Speech of the Egyptian Minister of the Interior, Major General Magdy Abdel Ghaffar, during the celebration of the 66th Police Day
Arabic Speech 9	Oct. 10, 1987	Hussein bin Talal, King of Jordan	Speech from the Throne at the opening of the fifth regular session of the tenth Jordanian Parliament
Arabic Speech 10	Sept. 30, 2013	Yusuf bin Alawi bin Abdullah, Sultanate of Oman's Minister Responsible for Foreign Affairs	The Sultanate of Oman's speech before the United Nations General Assembly at the sixty-eighth session of 2013

Arabic Speech 11	Dec. 13, 2017	Palestinian President Mahmoud Abbas	Speech of Palestinian President Mahmoud Abbas during the emergency summit of the Organization of Islamic Cooperation on Jerusalem in Istanbul, December 13, 2017
Arabic Speech 12	Jan. 11, 2020	Haitham bin Tariq bin Taimur, Sultan of Oman	Speech of His Majesty Sultan Haitham bin Tariq bin Taimur on the occasion of his accession to the throne on January 11, 2020
Arabic Speech 13	Oct. 24, 2018	Motazz Moussa, Prime Minister of Sudan	Review of the government plan and budget before the Sudanese Parliament
Arabic Speech 14	Jan. 20, 2019	Michel Aoun, President of Lebanon	Lebanese President's speech at the Beirut Economic Summit
Arabic Speech 15	Nov. 29, 2018	Bashar Al Asad, President of Syria	President Bashar Al-Assad's speech before the Syrian government
Arabic Speech 16	Mar. 9, 2015	Ibrahim Al-Jaafari, Iraqi Minister of Foreign Affairs	Speech of His Excellency Dr. Ibrahim Al-Jaafari, Iraqi Minister of Foreign Affairs, at the 143rd Regular Session of the Council of the League of Arab States
Arabic Speech 17	Feb. 25, 2019	Maeen Abdul Malik, Prime Minister of Yemen	Speech of Prime Minister Maeen Abdul Malik at the 40th session of the Human Rights Council, Geneva
Arabic Speech 18	Oct. 10, 2021	Hamad bin Isa Al Khalifa, King of Bahrain	The speech of His Majesty King Hamad bin Isa Al Khalifa at the opening of the fourth session of the fifth legislative term
Arabic Speech 19	Dec. 9, 2022	Aziz Akhannouch, Prime Minister of Morocco	Speech of the Moroccan King delivered by the Prime Minister at the Arab-Chinese Summit
Arabic Speech 20	Nov. 2, 2022	Mohamed Ould Ghazouani, President of Mauritania	Speech of President Mohamed Ould Ghazouani at the Arab Summit in Algeria
Arabic Speech 21	Jul. 25, 2021	Kais Saied, President of Tunisia	Tunisian President Kais Saied's speech and decisions to freeze parliament
Arabic Speech 22	Nov. 1, 2022	Ahmed Aboul Gheit, Secretary-General of the Arab League	Speech of the Secretary-General of the Arab League, Ahmed Aboul Gheit, at the opening of the 31st Arab Summit in Algeria
Arabic Speech 23	May 19, 2023	Mohamed Yunus al-Menfi, Chairman of the Presidential Council of Libya	Presidential Council Chairman's speech at the Arab Summit in Jeddah
Arabic Speech 24	Mar. 29, 2015	Ismail Omar Guelleh, President of Djibouti	Speech of Djibouti President Ismail Omar Guelleh at the Arab Summit in Sharm El-Sheikh

Editorial notes

Authorship contribution

Conceptualization: A. Al Zahran
Data collection: A. Al Zahran, R Jamoussi, E. Ali
Data analysis: A. Al Zahran, R. Jamoussi
Results and discussion: A. Al Zahran
Writing - review and editing: A. Al Zahran, R. Jamoussi, E. Ali

Research dataset
Not applicable.

Funding
Not applicable.

Image copyright
Not applicable.

Approval by ethics committee
Not applicable.

Conflicts of interest
The authors declare no conflict of interest.

Data availability statement
The raw data from this research may be made available by the first author upon request.

License

Publisher
Cadernos de Tradução is a publication of the Graduate Program in Translation Studies at the Federal University of Santa Catarina. The journal *Cadernos de Tradução* is hosted by the Portal de Periódicos UFSC. The ideas expressed in this paper are the responsibility of its authors and do not necessarily represent the views of the editors or the university.

Section editors
Andréia Guerini – Willian Moura

Style editors
Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

Article history
Received: 08-01-2025
Approved: 13-08-2025
Revised: 01-09-2025
Published: 10-2025