SKASE Journal of Translation and Interpretation, 2023; 16(2): 23–43
doi: 10.33542/JTI2023-2-3

Into B or not into B? The limited impact of interpreting direction on target text fluency and complexity

Marta Kajzer-Wietrzny, Adam Mickiewicz University
Agnieszka Chmiel, Adam Mickiewicz University

Abstract
Understood as a language rarely learned by non-native speakers (Whyatt & Pavlović 2021), Polish might be considered a language of limited diffusion. As a result, interpreters frequently work both into and from Polish. This stands in stark contrast with the approach to interpreting directionality preferred in many international organizations and representatives of the prominent Paris School, according to which retour (A-B) interpreting results in lower quality than B-A interpreting (Seleskovitch & Lederer 1989). As directionality is "one of the oldest issues discussed in the translation and interpreting community" (Gile 2005), it has managed to attract a number of scholars supporting either the 'into B' or the 'not-into B' approach. This paper aims to contribute to this long debate by focusing on the analysis of fluency and target text complexity as modulated by interpreting directionality. In order to see whether directionality influences interpreting quality, we identified potential differences in interpretations provided by the same interpreters working into A (Polish) or into B (English) and provided into English as either A language (traditional interpreting direction) or B language (retour interpreting). We used three subcorpora from PINC (Polish Interpreting Corpus) (Chmiel et al. 2023) and analysed a set of target text fluency and complexity measures. Our analysis showed mixed results when we compared the same interpreters' performance in two directions: fluency was higher in interpreting into B based on four measures (shorter silent pauses, fewer and shorter filled pauses, higher speech proportion) but lower based on three measures (more numerous silent pauses, shorter runs, lower articulation rate). When comparing interpreting into English as A or B language, we found lower fluency in interpreting into B based on five measures (longer silent pauses, more filled pauses, shorter runs, lower articulation rate and speech proportion) and no difference in the number of silent pauses and mean length of filled pauses. We also found no difference in target text complexity based on all measures with the exception of dependency distance: retour interpretations were more syntactically complex than interpretations into A language in the case of speeches delivered impromptu. Overall, our study shows that production fluency and complexity in retour interpreting are not consistently lower as compared to interpreting into A language. The findings thus weaken the case for interpreting into one's native language only.

Keywords*: simultaneous interpreting, directionality, fluency, complexity, corpus analysis*

1. Introduction

Conference interpreters provide language services at international meetings, where participants cannot communicate using a common language. In most such mediated events, interpreters are

native speakers of one language and non-native speakers of the other language in the communicative exchange. It is very rare that an interpreter has been raised in a bilingual culture and has become a native speaker of both. This means that one of the languages of the exchange will always be foreign for such an interpreter and the interpreting industry needs to accept this asymmetry. Now, whether the interpreter should interpret into their foreign language or not is a contentious matter.

Interpreting into a foreign language was pronounced as inferior and of lower quality than interpreting into one's mother tongue by the Paris School in the previous century (Seleskovitch & Lederer 1989). This approach is still reflected in many Western European countries, where languages like English, French or German are official. And yet, in Poland, just like in other countries of Central and Eastern Europe, and many other around the world, retour interpreting, i.e. interpreting out of the interpreter's native language (A language) into a foreign language (B language) is a common practice. A question immediately arises whether retour interpreting is really so much different from interpreting into the native tongue that it should be discouraged in professional practice and, in the context of international organizations, limited only to the cases where interpreters into A are not available.

The aim of this paper is thus to investigate in a quantitative way whether interpreting into B language is visibly inferior to interpreting into A. This will be done for parameters relating to its paralinguistic and linguistic form, i.e. fluency and text complexity. The findings might prove constructive to interpreting practitioners, teachers, students, as well as to conference organizers.

The concept of retour interpreting is closely related to the problem of language diffusion. Following Whyatt & Pavlovic (2021), we understand language of low diffusion (LLD) as a language whose spread outside the speech community is limited and the number of L2 speakers of such a language is modest. With a low number of L2 speakers, interpreting and translation into a given LLD needs to be performed by the L1 speaker of that language into a foreign language, whose diffusion is unlimited. This is precisely the case of English and Polish. The number of native speakers of English-speaking Polish is exceptionally small while the number of native Polish speakers speaking English is much larger. As a result, interpretations into Polish and other LLDs, even in international institutions, which favour interpreting into the mother tongue, are mostly but not entirely retour interpretations. In this paper we set to examine the features of retour interpreting based on authentic data from the European Parliament plenary sessions collected in the Polish Interpreting Corpus (PINC) focusing on fluency and target text complexity. In particular, we will be examining whether the same interpreters are equally fluent while interpreting simultaneously into the native and a foreign language. Looking from a different angle, we will also analyse whether different interpreters working into English as foreign language (B language) and as a native tongue (A language) produce equally fluent and equally complex interpretations. By investigating whether and how directionality influences the product of interpreting we want to revisit existing accounts regarding the potentially inferior status of retour interpreting, thus contributing to answering the question: into B or not into B.

The paper starts with a discussion of the key outcomes of the long tradition of research on directionality in interpreting. As we attempt to tap into quality by looking quantitatively at interpreting fluency and the complexity of the target language, we provide the readers with the background information about the measures used in investigations of fluency and complexity and outcomes of studies that serve as a basis for the preliminary assumptions of our own investigation. This is followed by the presentation of the study, its results and discussion.

2. Directionality in interpreting practice and interpreting research

Interpreters might have various languages in their combination. According to AIIC, the most prestigious professional association of conference interpreters, interpreters' working languages can be classified as follows: A, i.e. the interpreter's native language (or equivalent) "into which the interpreter works from all her or his other language; B: A language other than the interpreter's native language, of which she or he has a perfect command and into which she or he works from one or more of her or his other languages [...] C: Languages, of which the interpreter has a complete understanding and from which she or he works" (AIIC 2022). According to Öztürk (2020), interpreters working for international organizations such as the EU, UN or NATO are expected to interpret mostly into their native tongue as this direction is considered superior.

This attitude is not new, translation and interpreting into a foreign language was frowned upon in the 20th century in line with the belief that "only a native speaker or native signer of the target language – defined as someone born into a certain language community and having acquired its native language since birth – has the appropriate lingua-cultural competence to guarantee high-quality performance into that language" (Apfelthaler 2019: 152). Similar was the dominant perspective on directionality in Western Europe, highly influenced by the Paris School, with Seleskovitch & Lederer (1989) claiming that quality interpreting can only be assured when interpreters work into their mother tongue. However, retour interpreting is common in some Western European countries, e.g. Italy (Monti et al. 2005) or Poland. In countries with languages of low and limited diffusion, such as Slovenia, Hungary, Finland, Denmark and Turkey (Öztürk 2020), retour (A-B) interpreting is a regular practice. A survey carried out in Croatia (Pavlovic 2007) shows that not only over 70% or interpreters and translators regularly work into their foreign language, but also that one third of the surveyed population even prefers this direction and one third has no preference. According to Lim (2005), 33% of Korean interpreters complete half of the assignments into a foreign language, compared to 18% of AIIC interpreters, who report doing so.

Interpreting scholars report on directionality while investigating interpreting from the point of view of reception, cognitive effort, strategies applied, quality and modulating effects of background knowledge.

As regards the perception of interpretations into a foreign language by users of interpreting services, conference participants do not always distinguish between retour and B-A interpreting and retour is "on the whole appreciated and evaluated as being as good as SI into A" (Donovan 2002: 8).

Contrasting reports may be found in interpreting literature on directionality. There are a number of studies suggesting that interpreting into a foreign language is more cognitively taxing (e.g. Kurz 1993, Donovan 2002) and such presumed higher cognitive load might negatively impact the quality of interpreting. Kalina (2005: 42), however, points to the need to discuss directionality and quality not in general but rather in the context of "languages involved, type of conference, groups of participants for whom interpretation is intended, nationalities and cultural backgrounds of speakers, language distribution etc.". Moreover, quality can be looked at from different angles. In general, Denissenko (1989) found the quality of A-B interpretations higher than that of B-A interpreting. Some scholars link the improvement in quality in retour interpretation to an increase in comprehension (Chmiel 2016), which might lower the risk of erroneous interpretation. The perception depends on

operationalisation, however. For example, according to Öztürk (2020), interpreting quality operationalized with reference to meaning, target language use and delivery is in fact higher in interpretation into A, but the completeness of semantic content, i.e. propositional accuracy of the renditions into B language, was found to be more accurate. Retour interpreting might thus be easier due to facilitated comprehension in the native language and, simultaneously, more difficult due to more difficult production in the foreign language.

Processing in retour interpreting might differ from that in B-A interpreting. Bartłomiejczyk (2006) points to the fact that directionality may be the main reason behind differences in the use of interpreting strategies in Polish-English (A-B) versus English-Polish (B-A) interpreting, but she also suggests a possible impact of the language pair characteristics. On the other hand, Öztürk (2020) has not recorded any significant differences in the preferences for specific interpreting strategies across both directions. Directionality also seems to affect explicitation patterns to some extent, but the outcomes are not fully consistent. Tang (2018), for example, found that explicitation in retour interpreting is based on restructuring and paraphrasing information while interpreting into one's mother tongue triggered explicitation through addition of information. Explicitation in both directions would hence be differently motivated. To make the pattern of results even less consistent, Gumul & Bartłomiejczyk (2022) found the interpreters' explicitation behaviour not to be influenced by directionality in a systematic way.

Directionality in interpreting might also be modulated by the interpreter's background knowledge. The outcomes of a pilot study on the interaction between familiarity of the topic and interpreting quality suggest that better knowledge of the topic of interpreting is more beneficial for retour interpreting than for B-A interpreting (Dose 2017). This finding might be important in the context of retour interpretation in the EU institutions, where staff interpreters are more conversant with the discussed topics and might stand better chances of providing quality interpretation into a language that is foreign to them as compared to freelance interpreters with limited exposure to specific specialised topics.

As regards corpus studies focusing in particular on the interpretations into and from a mother tongue, there seems to be a terminological overlap. When introducing the European Parliament Corpus (EPIC), Monti et al. (2005) described it as a tool to investigate directionality understood in two different ways. One of them was indeed defined as the contrast between the interpretation into A vs. into B. The other definition, which has dominated the default understanding of the term "directionality" in corpus interpreting studies, refers to language combination, i.e. specific language pair e.g. Italian-English and the direction of language transfer Italian to English versus English to Italian. Hence, the majority of corpus studies referring to directionality, in fact, report on the characteristics of interpretations in specific language pair(s). Moreover, the vast majority of corpus studies on interpreting is carried out on the so-called "big" languages, frequently on data originating from the European Parliament or other international organizations, which favour interpretations into the mother tongue. With such language regimes it is more natural to focus on interpretations into L1 or L2 separately.

In the current paper, we return to the original dual understanding of directionality by Monti et al. (2005) as we will be examining Polish-English interpretations performed as interpreting into the mother tongue by interpreters from the English booth and performed as interpreting into B by interpreters from the Polish booth. Also, we will be examining Polish-English (A-B) and English-Polish (B-A) interpretations performed by interpreters from the Polish booth.

3. Text fluency in interpreting and directionality as one of its modulating factors

Fluency, understood after Rennert (2010) as a prosodic feature of speech based on such temporal variables as pauses, hesitations, audible breathing and speech rate, has been considered a relevant factor in perceiving and assessing interpreting. Surveys among conference interpreters have consistently shown the importance of fluency for interpreting quality: fluency ranked sixth among quality criteria in a study by Bühler (1986), fifth in the study by Zwischenberger (2010) and third in a global survey by Pöchhacker (2012). Users of simultaneous interpreting have also been found to consider fluency important: it ranked fifth in a study involving legal experts (Pradas Macías 2006) and in a survey among politicians, medical doctors and engineers (Kurz 1993). Fluency is thus featured in interpreting assessment criteria (Han et al. 2020) and students tend to rate more fluent interpretations as more accurate (Rennert, 2010).

Some studies have looked closer into the link between fluency parameters, perceived fluency and perceived accuracy. Studies featuring consecutive interpreting performed by students found that interpreting fluency and accuracy is perceived as higher if interpreting includes fewer filled pauses and shorter silent and filled pauses and when the articulation rate understood as the average speed of utterance without pauses is higher (Yu & van Heuven 2017). In another study focusing on consecutive interpretations by trainees (Han et al., 2020), perceived fluency of both A-B and B-A interpreting was shown to correlate with the duration of silent pauses, speech proportion, i.e. a ratio of articulation time (i.e. not including pauses) to speech duration, mean length of runs (understood as segments of speech uninterrupted by silent pauses) and speech rate. Taken together, these studies show a clear association made in the listeners' minds between fluency and accuracy.

Some experimental and corpus-based studies focusing on fluency parameters compare source and target texts, pinpointing differences in fluency between the two (Cecot 2001; Dayter 2021). Source texts are either regular spontaneously produced utterances or read out texts. In the latter, such fluency parameters as pauses can be either used as rhetorical devices or become manifestations of speech production difficulties (Tóth 2013). These difficulties might include searching for the right expression befitting an intended meaning. In simultaneous interpreting, pauses can arise for reasons not only related to production (looking for translation equivalents), but also related to comprehension and structural reformulation (Bartłomiejczyk 2006). According to Pöchhacker (2003), silent and filled pauses stem from limited planning characteristic for the specific type of oral production that is involved in interpreting. Therefore, in view of the speaker-induced constraints relevant to interpreting, it is no wonder that such studies consistently show lower fluency in target texts than source texts (Cecot, 2001). An interesting pattern is that interpreters produce fewer but longer silent pauses than speakers (Christodoulides, 2013; Wang & Li, 2015). Interpreters may use such longer pauses for information processing. A preliminary analysis of PINC data in which Polish source texts were compared to interpretations into English as a B language (Chmiel et al. 2022) also showed that interpreters produced shorter runs, or uninterrupted flows of utterance, than speakers. This finding indicates that interpretations are usually more fragmented due to processing constraints.

Interpreting fluency has been found to be modulated by many factors, including the interpreter's experience (Tóth 2013), sound quality (Piccaluga et al. 2005) and – with a special

relevance to the current study – directionality. Below we review studies that look at the effect of interpreting direction on fluency measures.

A-B interpretations are considered to be less fluent than B-A interpretations because production in the interpreter's B language induces more cognitive load and this cognitive difficulty is reflected in various fluency parameters, including filled pauses (Setton 1999). Beginner students, advanced trainees and professional interpreters were asked to interpret consecutively in both directions in a study conducted by Mead (2005). The mean duration of filled and silent pauses was greater in retour interpreting for all groups of participants, suggesting a more fragmented delivery into the B language. Filled pauses were much longer in retour interpreting than in B-A interpreting by students, which was interpreted as problems with production in retour. Mead also collected retrospective data from his participants and found that formulation was the most frequently offered explanation for the pauses produced with no directionality effect, i.e. pauses were triggered by similar causes irrespective of the interpreting direction. These results are partially in line with a smaller-scale study of four simultaneous interpreters (Piccaluga et al. 2005). In this case, retour interpreting was characterised by more frequent but shorter pauses. In another study, Lin et al. (2018) looked at how directionality influences fluency of interpreting trainees with Chinese as A language and English as B language. As expected, students produced less frequent interruptions (silent pauses) and hesitations (filled pauses) when interpreting into their native language.

The above review indicates that fluency is investigated in relation to directionality mainly in experimental studies that show a rather consistent effect: retour interpreting is less fluent and more fragmented than interpreting into the interpreter's A language. However, we have to bear in mind that these studies are either small-scale or based on data coming from trainees. To date, to the best of our knowledge, no large volume of data coming from professional interpreters has been analyzed to conclude about the link between fluency and directionality. Our study aims to fill this gap by looking at how directionality modulates fluency in interpreting performed by professional interpreters.

4. Text complexity in interpreting and directionality as one of its modulating factors

Another level at which interpretations can be analysed involves text complexity. These investigations are frequently made within the framework of Translation Universals, i.e. translation specific features which are believed to reflect cognitive constraints characteristic of the task (Baker 1993). This strand of research has recently been extended to constrained communication (Kotze 2020), which focuses on constrained varieties (translation/ interpreting among them) probabilistically conditioned by five dimensions, two of them being language activation (monolingual vs. bilingual) and text production (independent/unmediated vs. dependent/mediated). Viewed from this perspective, interpretation into the mother tongue is cognitively constrained by two dimensions: dependent text production, in which a text is produced based on another text (the speaker's utterance) and bilingual activation because the interpreter hears the text in one language and renders it in another. In retour interpreting the latter constraint is even stronger as the target text needs to be rendered in a foreign language.

In many studies, mediated (i.e. translated and/or interpreted) texts have been compared to written or spoken unmediated texts originally produced in the target language. The general assumption, in line with the simplification hypothesis, is that the translated and interpreted texts

would be less complex, i.e. simplified when compared to unmediated texts, due to the additional cognitive costs related to language mediation and dependent text production.

In most cases the investigated parameters focus on the core lexical patterns set out in a study carried out by Laviosa (1998), replicated to a great extent on interpreted texts in different language pairs. In interpreting, these parameters usually include lexical density (a ratio of content words to all words), core vocabulary coverage (frequency of most frequent words) and some form of type-token ratio (TTR). It has been proven in some cases indeed, that complexity of the target text in interpreting is lower than complexity of spoken unmediated texts (e.g. Bernardini et al. 2016), but other studies (Russo et al. 2006; Dayter 2018) show that these tendencies are dependent on language pair and other factors (e.g. whether the source was originally read out or delivered impromptu, delivery rate of the original).

It turns out that complexity measures can also be indicative of the quality of interpreting. Ouyang et al. (2021) used shallow statistics, such as word count, word diversity, and hypernymy of verbs to predict quality assessments in an interpreting contest. They found that interpretations by better rated interpreters used more sophisticated language. In this context, it is thus worth investigating text complexity measures to see the potential modulatory effect of directionality using a large volume of naturalistic data from the bidirectional corpus, as we have undertaken in the present study.

Another lexical measure investigated in this context involves the way users combine words together, i.e. collocativity. Two measures tapping into different features of collocations have been explored so far to investigate how the use of collocations differs across mediated (interpreted) and unmediated (original) texts: T-score and MI-score (Mutual Information score). High T-score values typify collocations, which are highly frequent and high MI-score values are characteristic of texts with collocations that are not so frequent but their components are very strongly associated and are more difficult to retrieve (Ferraresi 2019). Previous studies show that mediated texts do not differ from independent written or oral text production in terms of the use of high T-score collocations but interpretations do make significantly less frequent use of high MI-score collocations than translations (Ferraresi & Milicevic 2017, Ferraresi 2019). What is highly relevant for the study reported here is that Ferraresi (2019) found speeches originally read out to be more collocational than those delivered impromptu and that collocational values of interpretations and of speeches delivered by non-natives do not differ significantly from each other but differ from the speeches delivered by natives.

Text complexity can be also investigated at the level of syntax. For example, Liang et al. (2017) compared mean dependency distance (MDD) in translation, simultaneous interpreting and consecutive interpreting. Dependency distance covers the number of elements between syntactically dependent elements. Greater dependency distance indicates greater syntactic complexity, which is potentially more challenging to process. Current evidence points to the greatest syntactic complexity of translated texts, followed by simultaneous interpreting and by consecutive interpreting, where dependency distance is the shortest. The authors suggest that translation has the lowest risk of cognitive saturation but also that following the source text sentence by sentence in simultaneous interpreting might influence MDD in this mode more than in consecutive interpreting. More importantly for the study reported in this paper, Jiang & Jiang (2020) looked at the relation between maximum dependency distance and interpreter's disfluencies and found that long maximum dependency distance sentences are associated with more disfluencies than short sentences.

The only corpus study that compares complexity measures across interpretation into A and into B is by Huang et al. (2023) with a focus on formulaic structures. The study examined

formulaic sequences in the form variations in sets of fixed phraseological frames differing in one word (the so-called p-frames) among original English speeches, Chinese-English interpretations into A and Chinese-English interpretations into B in a comparable corpus. The results showed that task complexity and directionality contributed to the increase of cognitive load, which positively correlated with the number of p-frames pointing to the use of a more formulaic language. Retour interpreting used the greatest number of formulaic structures, while original English texts relied on them less. Therefore, the study concludes that the more cognitive load involved in interpreting, the greater the reliance on formulaic sequences – the interpreter simply chooses the most obvious and easy-to-process solutions to minimise effort.

The authors observed a similar level of fixedness of formulaic structures in both interpreting directions, which means that the proportion of structures typical and untypical for the target language used in both cases was similar. However, the direction of interpreting turned out to have an impact on the formulaic structures. When interpreting into B, interpreters favoured expressions with content words (as opposed to expressions based on function words), a finding that suggests a tendency towards greater precision (Huang et al. 2023), while interpreters working into A chose expressions based on function words, which suggests greater dilution of information.

Taken together, these studies suggest a tendency of mediated language to be less complex. This tendency is less stable in interpreting, the spoken mediated variety, and seems to depend on the analysed language pair. Meanwhile, studies that specifically compare complexity of target texts produced in interpreting into B language and into A are rare and there are reasons to believe that the higher cognitive load caused by multiplication of cognitive constraints in retour interpreting might shape the form of the target text. This is one of the research gaps that we are going to address in this paper.

5. The present study

The aim of the present study is to investigate retour interpreting on the basis of naturalistic data from professional interpreters. We want to examine parameters of authentic interpretations in terms of fluency and text complexity to see if we can find evidence for inferior performance in retour interpreting as compared to B-A interpreting. For that purpose, we take advantage of the unique features offered by PINC, the Polish Interpreting Corpus (Chmiel et al. 2023) that comprises authentic interpretations from the European Parliament plenary sittings. In the first step, we will identify potential differences in interpretations provided by the same interpreters working into A (Polish) or into B (English) and in the second, we will look at interpretations provided by different interpreters into English as either A language (traditional interpreting direction) or B language (retour interpreting). Thus, we will be able to use the within-subject design to compare the performance of the same interpreters from the Polish booth who either interpret from English into Polish or perform retour interpreting from Polish into English. We will also use the between-subject design to compare Polish-English interpreting performed by different interpreters working into English as their A or B language. As this is a corpus-based study, we will use multiple dependent variables drawn from the plethora of PINC metadata. We will be able to use interpretation fluency measures in both comparisons as these seem rather universal for the languages in question. However, evidence exists showing that text complexity measures are language specific (Strömqvist et al. 2002) and thus we can only use such measures when comparing interpretations into the same

language (i.e. English as language A or B). Therefore, the between-subject analysis will include both fluency and complexity measures.

We predict that interpretations into B will be less fluent, which will be manifested through more numerous and longer pauses, lower articulation rate, speech proportion and shorter runs following previous findings from experimental (Mead 2005; Han et al. 2020, Piccaluga et al. 2005; Lin et al. 2018) and other studies (Setton 1999). As regards text complexity measures, we hypothesize that as the effect of bilingual constraint dimension (Kotze 2020) on retour interpretations is amplified by non-native production, these interpretations will be less complex. In other words, we envisage that retour interpreting will display lower Mean Segmental Type-Token Ratio (MSTTR), greater list head coverage, lower lexical density, lower MI-score and higher T-score, lower Fog index, and lower dependency distance as compared to B-A interpreting. All the dependent measures are explained later in this section.

5.1.Dataset: Polish Interpreting Corpus (PINC)

PINC comprises transcripts of speeches and interpretations annotated with rich paralinguistic information. In the study reported here we used three PINC subcorpora. Two of them were previously described (Chmiel et al. 2022; 2023): **PL-SI-A,** interpretations into Polish (as a mother tongue) and **EN-SI-B,** interpretations into English (retour made by Polish interpreters). The third was a recently compiled **EN-SI-A** subcorpus consisting of interpretations into English as a mother tongue. As visible in Table 1, the latter is much smaller, which also reflects the LLD status of Polish, as Polish-English interpreting into A at the EP is relatively rare.

Table 1: Number of running words in PINC subcorpora

Corpus	Delivery	Wordcount
PL-SI-A	Impromptu	17496
	Read	19153
EN-SI-A	Impromptu	5469
	Read	5161
EN-SI-B	Impromptu	21452
	Read	19453

The texts in PINC range between 100 and 500 words, with the mean text length of 204. PINC is annotated with rich metadata regarding each speech in the corpus including details about, for instance, the original speaker, the debate, the interpreter, the mode of delivery, delivery rate.

Following the EPIC corpus (Monti et al. 2005), the mode of delivery of the source speech is encoded, but only speeches originally read out or delivered impromptu are included.

There are altogether 39 Polish native interpreters in the corpus, interpreting both into their mother tongue (A) and their foreign language (B). The native English interpretations are carried out only by three interpreters as Polish is not a popular C language among native English interpreters working for the EU. The voices of individual interpreters were first identified by humans and then in an automated procedure (see Koržinek & Chmiel 2021).

As speech-to-text alignment was crucial for the project, PINC data have been also meticulously transcribed and timestamped. We employed an automatic speech recognition system followed by manual correction (see Chmiel et al. 2022). As a result, we have achieved detailed sound-to-text word-level alignment, which in practice means that all words, pauses and disfluencies are orthographically transcribed, timestamped and available for analysis. A description of the speech-text alignment in PINC is presented in Koržinek (2020).

5.2.Operationalisations of fluency and complexity measures

Thanks to extensive processing of audio recordings and transcriptions (Chmiel et al. 2022, Koržinek & Chmiel 2021), PINC offers a rich selection of metadata that we were able to use in the analyses. We focused on text fluency measures and text complexity measures. Seven fluency measures were used. The first four were related to pauses. We identified silent pauses as periods of silence longer than 250 ms following Han et al. (2020) and Mead (2005), while there was no cut-off point for filled pauses, in accordance with Plevoets & Defrancq (2016). We calculated the number of silent and filled pauses per each interpretation from the corpus and normalised the number per minute. We also measured the duration of all pauses and calculated the mean length of silent and filled pauses per interpretation. We then calculated the mean length of runs expressed in syllables, understood as segments of speech uninterrupted by silent pauses (Han et al. 2020). Articulation rate was calculated as the average speed of utterance without pauses (Christodoulides, 2013), while speech proportion was understood as a ratio of articulation time (i.e. not including pauses) to speech duration.

Following previous studies reviewed above, we assumed that fluency would positively correlate with run length, articulation rate and speech proportion and negatively correlate with the number and duration of pauses. In other words, higher fluency was operationalised as less numerous and shorter silent and filled pauses, longer runs, higher articulation rate and higher speech proportion. More fluent interpreters would thus use fewer and shorter pauses, produce longer utterances without pausing, produce utterances faster and use more time from the interpretation for actual oral production rather than pausing.

We used a number of measures to investigate lexical complexity. We started with a crude measure of lexical richness, i.e. Mean Segmental Type-Token Ratio (MSTTR) calculated using the Quanteda package (Benoit et al. 2018) in R (R core team 2021). The algorithm for MSTTR divides tokens into segments of a specific size, which in this case was 100, and type-token-ratio is calculated for each segment. Higher MSTTR is interpreted as an indication of higher lexical richness. Next, as other studies on interpreting (Bernardini et al. 2016, Dayter 2018), we followed Laviosa (1998) and used lexical density, i.e. the proportion of content words to all words in the text, to measure how informative interpretations are. Higher lexical density is interpreted as an indication of greater informativeness. For the same reason, we measured repetitiveness with listhead coverage, i.e. the proportion of the corpus covered by the 100 most frequent words. Greater lexical coverage is interpreted as an indication of greater tendency to repetitiveness. As regards phraseology, the approach in the reported analysis was inspired by the study by Ferraresi (2019). We measured indices of collocativity to tap into the

two various types of collocations: the most frequent ones (with high T-score) and the less frequent but strongly associated ones (with high MI-score). Just as Ferraresi, we first extracted the collocations from PINC, focusing on the selected part-of-speech patterns: adjective + noun, noun + noun, verb + noun, noun + verb. Next, we calculated T- and MI-scores of individual collocations based on frequencies of a large reference corpus (ukWaC; Baroni et al. 2009) as such a method reflects the use of collocations in general English (Ferraresi 2019). We then calculated mean T-scores and mean MI-scores of all the investigated collocations for all texts in PINC. This allowed us to see which interpretations rely more on frequent collocations as opposed to less frequent but more associated ones.

We also used yet another crude complexity measure, which taps both into lexis and syntax from the Quanteda package, namely Gunning's Fog Index (Gunning 1952). This index takes into account such a lexical measure as word length and the syntactic measure of average sentence length (utterances in PINC are transcribed as sentences to facilitate the use of NLP tools). Finally, we operationalized syntactic complexity with mean dependency distance. Having annotated the corpus with UDPipe package (Wijfels et al. 2019), we calculated dependency distance, following Jiang & Jiang (2020), by subtracting the position of a dependent element from the position of the governing element in a sentence.

6. Data analysis and results

We performed two sets of analyses. We first compared interpretations performed by the same interpreters in a within-subject design in two directions: from English into Polish as their A language and from Polish into English as their B language (retour interpreting). This analysis involved the above-described dependent variables operationalising text fluency. We then compared interpretations made by different groups of interpreters always from Polish (either A or C language) into English (either B or A language). This between-groups design allowed us to compare interpretations into the same language (English) with a different status (A or C). Because we analysed English as the target language, this analysis included not only text fluency but also text complexity measures. All analyses were performed as linear mixed effects models fitted in R (R Core Team, 2021). Dependent variables were log-transformed or transformed with the transformTukey function from rcompanion package (Mangiafico 2023) if required to arrive at a normal distribution. We used sliding contrasts with contr.sdif function from MASS package (Venables & Ripley, 2002) for contrast coding, interpreter identification data as our random factor and we added a term for the mode of delivery to control for the potential effect of impromptu or read out speeches on fluency. Mode of delivery will only be reported below if significant.

6.1. B-A (English-Polish) and A-B (Polish-English) interpreting performed by the same interpreters

The first model included the number of silent pauses per minute as a dependent variable. We found a statistically significant effect of directionality (b=-1.33, SE=.30, t=-4.37, p<.001), mode of delivery (b=-.58, SE=.26, t=-2.16, p=.031), and an interaction (b=1.30, SE=.53, t=2.42, p=.016). In line with our predictions, retour interpreting was characterised by a higher number of silent pauses (M=10.35, SD=3.40) as compared to B-A interpreting (M=9.34, SD=3.55), especially for the impromptu speeches. Surprisingly, silent pauses were longer in

B-A interpreting (M=709 ms, SD=348) than in A-B interpreting (M=626 ms, SD=263),(b=.08, SE=.04, t=2.20, p=.028) and there was a lot of variation for both directions, as seen by high standard deviation values. Again, contrary to our predictions, filled pauses were more numerous in B-A interpreting (M=8.74, SD=4.82) than in retour interpreting (M=7.37, SD=5.20), (b=1.94, SE-.43, t=4.50, p<.001). Mode of delivery also influenced the number of filled pauses, with more numerous filled pauses in interpretations of read out speeches in both directions (b=.96, SE.38, t=2.52, p=.012). Filled pauses were also longer in interpreting into A (M=804 ms, SD=407) than in retour interpreting (M=722 ms, SD=273), (b=70.26, SE=34.13, t=2.06, p=.04).

We then looked into more complex fluency measures. We found a directionality effect in our analysis of the mean length of runs (b=8.26, SE=1.02, t=8.11, p<.001). As expected, runs were longer in B-A interpreting (M=27.86, SD=11.99) than in retour interpreting (M=20.20, SD=9.25), attesting to the more fluent and less fragmented delivery in the former. Our analysis also revealed a directionality effect (b=74.95, SE=3.00, t=24.91, p<.001) and an interaction between direction and mode of delivery (b=-13.37, SE=5.25, t=-2.54, p=.01) for articulation rate. Articulation rate was higher in B-A interpreting (M=318, SD=41) than in A-B interpreting (M=248, SD=42) and the interaction was driven mainly by a higher articulation rate in B-A interpreting for impromptu rather than read out speeches. We also found a directionality effect for speech proportion (b=-.02, SE=.006, t=-3.56, p=.0004), with a higher value for retour interpreting (M=.80, SD=.07) than for B-A interpreting (M=.77, SD=.08).

Table 2 includes a summary of all fluency measures used in the comparison of B-A (English-Polish) and A-B (Polish-English) interpretations performed by the same interpreters. We see that the fluency of retour is higher for four measures and lower for three measures.

Table 2: Fluency measures and their meaning in the comparison of retour interpreting to interpreting into A language

Fluency measure	Value of the measure in retour interpreting (as compared to interpreting into A)	Fluency of retour interpreting (as compared to interpreting into A)
Number of silent pauses per minute	higher	lower
Length of silent pauses	lower	higher
Number of filled pauses per minute	lower	higher
Length of filled pauses	lower	higher
Mean length of run	lower	lower
Articulation rate	lower	lower
Speech proportion	higher	higher

6.2. Polish-English interpreting performed as interpreting into A and interpreting into B

We first looked into silent and filled pauses, both their number normalised per minute of speech and their mean length. We found no statistically significant effects in the analysis of the number of silent pauses per minute: retour interpreting included a similar number of silent pauses (M=10.36, SD=3.14) as into-A interpreting (M=11.24, SD=3.40). The analysis of the mean length of silent pauses showed that pauses were longer in retour interpreting (M=626 ms,

SD=264) than interpreting into A (M=454 ms, SD=122), with a much higher variation in the former (b=0.25, SD=.10, t=2.44, p=.026). We then looked at similar measures for filled pauses. This time, we found a mode of delivery effect (b=-1.49, SE=.55, t=-2.71, p=.007) and an interaction between the mode of delivery and direction (b=4.99, SE=1.1, t-4.53, p<.0001), which was driven mainly by a huge difference in the number of filled pauses in interpretations of read out speeches (M=1.27 for interpreting into A and M=8.39 for interpreting into B), thus showing retour interpreting to be less fluent than interpreting into A. We found no effects for the mean length of filled pauses.

We then fitted three more models for three more complex fluency measures: runs, articulation rate and speech proportion. We found directionality effect in the mean length of runs (b=-5.96, SE=2.47, t=-2.41, p=.03) with longer runs in interpreting into A (M=26.01, SD=8.46) than in retour interpreting (M=20.20, SD=9.25). The analysis of articulation rate brought a statistically significant effect of direction (b=-79.55, SE=5.72, t=-13.89, p<.001), mode of delivery (b=15.15, SE=3.66, t=4.14, p<.001) and an interaction of the two (b=-24.75, SE=7.31, t=-3.38, p=.0008). It turned out that articulation rate was higher in interpreting into A (M=327, SD=33) than in retour (M=248, SD=24) and the interaction was driven mainly by much higher articulation rate in B-A interpreting of read out speeches as compared to impromptu speeches. We also found directionality and mode of delivery effects in the analysis of speech proportion. Speech proportion was higher in B-A interpreting (M=.88, SD=.04) than in retour (M=.80, SD=.07), (b=-.06, SE=.02, t=-2.41, p=.025) and higher in interpreting read out speeches (M=.82, SD=.08) than impromptu speeches (M=.81, SD=.07), (b=.02, SE=.007, t=.2.20, p=.02).

Another set of analyses focused on six target text complexity measures. We fitted linear mixed effects models in a similar way. We found no directionality or mode of delivery effects for lexical richness, listhead coverage and T-score. We found a mode of delivery effect for lexical density (b=.01, SE=.004, t=4.38, p<.0001) with higher lexical density for interpretations of read out speeches (M=97.05, SD=5.28) as compared to impromptu speeches (M=96.57, SD=5.76). A similar effect was found for the Fog index (b=.92, SE=.35, t=2.60, p<.009) with a higher value for interpretations of read out speeches (M=14.46, SD=2.44) than impromptu speeches (M=13.69, SD=2.84). Another mode of delivery effect was found for MI-score (b=7.06, SE=2.31, t=3.05, p=.002) with higher MI-score for interpreting read out speeches (M=32.70, SD=16.50) than impromptu speeches (M=26.30, SD=15.44). Our analysis of dependency distance revealed a statistically significant interaction between directionality and the mode of delivery (b=-.15, SE=.07, t=-2.12, p=.03) driven by the difference between two interpreting directions when interpreting impromptu speeches (M=2.54, SD=.21 for interpreting into A and M=2.70, SD=.27 for interpreting into B).

Table 3 presents a summary of all the fluency and complexity measures used in the comparison of Polish-English interpretations performed by the English booth (into A) and by the Polish booth (retour into B). We see that the fluency of retour is lower for five out of seven measures and similar for two measures. Text complexity produced in retour is higher for one measure and similar to interpreting into A for six measures.

Table 3: Fluency and complexity measures and their meaning in the comparison of retour interpreting to interpreting into A language.

Fluency measure	Value of the measure in retour interpreting (as compared to interpreting into A)	Fluency of retour interpreting (as compared to interpreting into A)
Number of silent pauses per minute	similar	similar
Length of silent pauses	higher	lower
Number of filled pauses per minute	higher in interpreting read out speeches	lower
Length of filled pauses	similar	similar
Mean length of run	lower	lower
Articulation rate	lower	lower
Speech proportion	lower	lower
Complexity measure	**Value of the measure in retour interpreting (as compared to interpreting into A)**	**Complexity of retour interpreting (as compared to interpreting into A)**
MSTTR	similar	similar
Listhead coverage	similar	similar
Lexical density	similar	similar
Dependency distance	higher for impromptu speeches	higher
Fog index	similar	similar
MI-score	similar	similar
T-score	similar	similar

7. Discussion

The aim of this study was to analyse text fluency and complexity parameters in authentic interpretations performed by experienced professional interpreters to see if and how directionality modulates these parameters. More specifically, we wanted to see if naturalistic data confirms the generally held view and results of some experimental and corpus-based studies about inferior interpreting performance into the B language. As both fluency and complexity have been found to be associated with interpreting quality (Yu & van Heuven 2017, Ouyang et al. 2021), we decided to use them in an analysis of directionality in interpreting, even though actual quality measures were not used. In studies based on larger volumes of data compiled into corpora, like ours, the application of more standard measures of quality is not feasible. This is why we decided to opt for the automatically derived measures that target specific components of interpreting quality, namely fluency and complexity.

7.1. Fluency of retour interpreting is not consistently lower than fluency in B-A interpreting

We first compared fluency measures for English-Polish (B-A) and Polish-English (A-B) interpretations performed by the same interpreters in a within-subject design. We predicted that retour interpretations will be less fluent. This prediction was confirmed only for three out of seven measures (number of silent pauses per minute, mean length of run and articulation rate). Contrary to our expectations, as many as four measures indicated retour superiority as regards fluency, i.e. interpreters produced shorter silent pauses, fewer and shorter filled pauses in retour interpreting and they also had a higher speech proportion, i.e. they used a larger proportion of the interpreting time for actual oral production. These findings are in line with Piccaluga et al. (2005) who also observed a pattern of more numerous but shorter silent pauses in simultaneous interpreting. However, they are at a variance with Mead (2005), who found longer filled and silent pauses in retour interpreting. This discrepancy might be explained by the task: Mead (2005) investigated consecutive interpreting which, due to its specificity and fewer immediate speaker-induced temporal constraints), might manifest a different fluency pattern than simultaneous interpreting. These results are also partially at a variance with Lin et al. (2018), who found more numerous both filled and silent pauses in retour simultaneous interpreting. The data from Lin et al. (2018) come from trainee interpreters - unlike professionals in our study, they produced more filled pauses in retour interpreting. While silent pauses (produced in a similar pattern by trainees in Lin et al.'s study and ours) might reflect comprehension load (Setton 1999), filled pauses suggest production load and this might have been greater for the B language for trainees as compared to professional interpreters. This might explain the difference in our findings.

Interpreters featured in our corpus produced fewer filled pauses and shorter filled and silent pauses in retour interpreting - if we applied the findings by Yu & van Heuven (2017) to these fluency patterns, we might conclude that retour interpretations could be perceived as more fluent and more accurate. The reverse would be true for articulation rate - it was lower in retour interpreting and, again according to Yu & van Heuven (2017) could lead to perceiving this interpretation as less fluent and less accurate than B-A interpreting. On the other hand, according to Han et al. (2020), length of silent pauses, speech proportion and run length correlate with fluency perception and our findings suggest that two out of three measures would point to the superiority of retour interpreting. As we can see, the pattern of results stemming from the first comparison in this study does not consistently point to higher fluency of B-A interpreting. In fact, many measures indicate the opposite. Let us discuss the other analysis before drawing general conclusions.

The second comparison pertained to Polish-English interpreting by different groups of interpreters with English as A or B language. This time the pattern of results was slightly different. Five measures pointed to the superiority of B-A interpreting and two measures indicated a similarity between the two directions. The findings matched the previous analysis in relation to two measures, mean length of run and articulation rate, both of which pointed to greater fluency of B-A interpreting. This was in line with our predictions and indicates more fragmented delivery in the B language. When compared to the previous analysis, the results from the current analysis showed reversed patterns for the length of silent pauses, number of filled pauses and speech proportion, thus in line with Mead (2005) and Lin et al. (2018) but contrary to Piccaluga et al. (2005). If explained following Setton's (1999) understanding of pauses, these results point to comprehension load resulting in longer silent pauses and production load resulting in more numerous filled pauses in retour interpreting. One has to bear

in mind that this analysis was performed in a between-groups design, i.e. it has less statistical power than the previous within-participant analysis since individual differences between Polish booth and English booth interpreters could not be controlled and could have skewed the results. As regards the number of silent pauses and the length of filled pauses, both groups performed similarly as directionality did not modulate either of these measures. More research is needed since our pattern of results is not consistent. However, what should be emphasized in light of our predictions, text fluency measures do not unanimously point at the superiority of B-A interpreting over retour, thus lending some empirical support in favour of retour interpreting.

7.2. Directionality has smaller impact on target text complexity than mode of delivery of the source speech

Looking at the texts of A-B and B-A interpreting through the lens of our text complexity measures, it is rather difficult to tell the two apart. In our models lexical richness, lexical density, listhead coverage, T-score, MI-score and Fog index seem to be immune to the effect of directionality. Because our corpora include both interpretations of speeches originally delivered impromptu and read out, we needed to control for this factor in our models and it proved to be significant for three of our measures, i.e. lexical density, MI-score and Fog index. Finally, one of the measures, namely mean dependency distance, turned out to highlight the difference between the texts of interpretations into B and into A but only for impromptu speeches. We found that when the original speeches are delivered impromptu, interpretations into B use more complex syntactic structures characterised by higher mean dependency distance than interpretations into A.

The first observation might not be that surprising. Interpreters at the European Parliament are recruited from the best professionals in the market and their language B must meet very high standards. Moreover, it has been found that directionality might also be modulated by the interpreter's background knowledge and that being conversant with the topic is more beneficial for retour interpreting than for interpreting into A (Dose 2017). In this context, one might suppose that being part of the European institutions, EP interpreters are highly familiar with the topics discussed at the plenaries, which helps them bring their performance to the native-like level. It is also not surprising that lexical density, MI-score and Fog index are sensitive to the impact of the mode of delivery of the original source more than to the impact of directionality. The differences between the spoken and written register are very pronounced and the effect of register usually takes over the subtle nuances of language mediation (e.g. Ivaska & Bernardini 2020). The final observation regarding the interaction between directionality and mode of delivery is less self-explanatory. More complex syntactic structures are typically characteristic of the more formal, written register. We can hypothesize that while interpreting an impromptu speech into a foreign language, interpreters, consciously or not, might attempt to make the target text sound more formal as research shows that non-native speakers prefer more formal linguistic options (see e.g. Alipour & Nooreddinmoosa 2018) and interpreters tend to make their target text sound more formal than unmediated speeches (Ivaska et al. 2022). This effect might be amplified by the fact that the language mediation involved is A-B interpreting, i.e. production in a foreign language. Other than that it seems that there are no strong differences between the linguistic features in the interpretations.

In general, our analysis shows that the fluency parameters are not consistently lower in retour interpreting and the investigated linguistic parameters for the interpretations into A and

into B are to a great extent similar. Both strands of our analysis therefore seem to diminish the strength of the arguments against retour interpreting.

8. Limitations

In the reported study we only focused on fluency and complexity measures and previous research shows that they can be associated with quality (Yu & van Heuven 2017, Ouyang et al. 2021). Future analyses of the kind could benefit from controlling more detailed aspects of the source speech to a greater extent. Moreover, without a detailed accuracy measure we cannot conclude with certainty that interpreting into one direction is similar or better than interpreting into the other direction. This is a limitation because an interpretation may be fluent and linguistically and syntactically complex, but still inaccurate. This study has been an attempt to fill the gap created by the lack of corpus-based studies using authentic data from professionals that would focus on directionality.

9. Conclusions

In this paper we attempted to show that even though interpreting into B has been viewed as inferior, there might be good reasons explaining why it has been practiced in many countries. First, there is no alternative when dealing with a language of limited diffusion. Second, as transpires from our analysis, text complexity parameters of interpretations in both directions are comparable. As regards fluency, it is shown to be either higher or lower in retour as compared to B-A interpreting, depending on the measure analysed. While our outcomes do not unanimously suggest that interpreting into B is superior, they certainly undermine some of the central arguments in the discussion against it. We are looking forward to the development of reliable automatic quality metrics that could help estimate propositional accuracy in parallel interpreting corpora, such as PINC. This is one of the avenues that future researchers could take to help determine the impact of directionality on interpreting quality, which is so vital both for the interpreter training and interpreting practice.

Funding information
This work was supported by the National Science Centre, Poland under Grant No. 2018/30/E/HS2/00035 awarded to Agnieszka Chmiel.

References

AIIC. 2022. Regulation governing admissions and language classification. Available at (https://aiic.org/document/10231/CACL_Regulations_2022_E&F_final.pdf) (Accessed 2023-10-02).

Alipour, Mohammad & Mona Nooreddinmoosa. 2018. Informality in Applied Linguistics Research Articles: Comparing Native and Non-Native Writings. *Eurasian Journal of Applied Linguistics*, 4 (2), 349–73.

Apfelthaler, Matthias. 2019. Directionality. In Baker, M. G. Saldanha (eds.). *Routledge Encyclopaedia of Translation Studies*, Routledge: London, 152-156.

Baker, Mona. 1993. Corpus Linguistics and Translation Studies: Implications and applications. In Baker,M. G. Francis & E. Tognini-Bonell (eds.). *Text and Technology: In honour of John Sinclair*, 233-250. Amsterdam/Philadelphia: John Benjamins.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. *The WaCky wide web: a collection of very large linguistically processed web-crawled corpora.* Language resources and evaluation, 43, 209–26.

Bartłomiejczyk, Magdalena. 2006. Strategies of simultaneous interpreting and directionality. *Interpreting*, 8 (2), 149–74.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller & Akitaka Matsuo. 2018. quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3 (30), 774–774.

Bernardini, Silvia, Adriano Ferraresi & Maja Miličević. 2016. From EPIC to EPTIC—Exploring simplification in interpreting and translation from an intermodal perspective. *Target*, 28 (1), 61–86.

Bühler, Hildegund. 1986. Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5 (4), 231–35.

Cecot, Michela. 2001. Pauses in simultaneous interpretation: A contrastive analysis of professional interpreters' performances. *The interpreters' newsletter*, 11, 63–85.

Chmiel, Agnieszka. 2016. Directionality and context effects in word translation tasks performed by conference interpreters. *Poznan Studies in Contemporary Linguistics*, 52 (2), 269–95.

Chmiel, Agnieszka, Janikowski, Przemysław, Koržinek, Danijel, Lijewska, Agnieszka, Kajzer-Wietrzny, Marta, Jakubowski, Dariusz & Plevoets, Koen. 2023. Lexical frequency modulates current cognitive load, but triggers no spillover effect in interpreting. *Perspectives*, 1–19.

Chmiel, Agnieszka, Danijel Korzinek, Marta Kajzer-Wietrzny, Przemysław Janikowski, Dariusz Jakubowski, and Dominika Polakowska. 2022. Fluency parameters in the Polish Interpreting Corpus (PINC). Mediated discourse at the European Parliament, 63.

Christodoulides, George. 2013. Prosodic features of simultaneous interpreting. In P. Mertens & A. C. Simon (eds.). *Proceedings of the Prosody-Discourse Interface Conference 2013* (IDP-2013), 33–37.

Dayter, Daria. 2018. Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN). *Forum*, 241–64.

Dayter, Daria. 2021. Variation in non-fluencies in a corpus of simultaneous interpreting vs. non-interpreted English. *Perspectives*, 29 (4), 489–506.

Denissenko, Jurij. 1989. Communicative and interpretative linguistics. In Gran, L.& J.M. Dodds (eds.). *The theoretical and practical aspects of teaching conference interpretation*, Udine, Campanotto Editore, 155–57.

Donovan, Clare. 2002. Survey of user expectations and needs. In Teaching simultaneous interpretation into a 'B' language: *EMCI Workshop proceedings, EMCI*, 2–11.

Dose, Stefanie. 2017. Assessing directionality in context. *Stellenbosch Papers in Linguistics*, 47, 67–87.

Ferraresi, Adriano. 2019. Collocations in contact: Exploring constrained varieties of English through corpora. *Textus*, 32 (1), 203–22.

Ferraresi, Adriano & Miličević, Maja. 2017. Phraseological patterns in interpreting and translation: Similar or different. In G. De Sutter, M.-A. Lefer, & I. Delaere, *Empirical translation studies: New methodological and theoretical traditions*, 157–82.

Folaron, Debbie. 2015. Introduction: Translation and minority, lesser-used and lesser-translated languages and cultures. *The journal of specialised translation*, 24, 16–27.

Gile, Daniel. 2005. Directionality in conference interpreting: A cognitive view. In: Rita Godijns and Michaël Hinderdael (eds.). Directionality in Interpreting: The 'Retour' or the Native? *Communication & Cognition*. 38(1/2), 9-26.

Gunning, Robert. 1952. *The Technique of Clear Writing*. New York: McGraw-Hill.

Gumul, Ewa & Magdalena Bartłomiejczyk. 2022. Interpreters' explicitating styles: A corpus study of material from the European Parliament. *Interpreting*, 24 (2), 163–91.

Han, Chao, Sijia Chen, Rongbo Fu & Qin Fan. 2020. Modeling the relationship between utterance fluency and raters' perceived fluency of consecutive interpreting. *Interpreting*, 22 (2), 211–37.

Huang, Dan Feng, Fang Li & Hang Guo. 2023. Chunking in simultaneous interpreting: the impact of task complexity and translation directionality on lexical bundles. *Frontiers in Psychology*, 14.

Ivaska, Ilmari & Bernardini, Silvia. 2020. Constrained language use in Finnish: A corpus-driven approach. *Nordic Journal of Linguistics*, 43(1),33-57.

Ilmari Ivaska, Adriano Ferraresi & Marta Kajzer-Wietrzny. 2022. Formality in mediated and non-mediated discourse: Bringing together human judgements and corpus-driven detection. In Kajzer-Wietrzny,M. Ferraresi,A. Ivaska,I. & Bernardini, S. (eds.), *Mediated discourse at the European Parliament: Empirical investigations*, 29–61. Berlin: Language Science Press.

Jiang, Xinlei & Yue Jiang. 2020. Effect of dependency distance of source text on disfluencies in interpreting. *Lingua*, 243, 102873.

Kalina, Sylvia. 2005. Quality in the Interpreting Process: What Can Be Measured and How?. *Communication and Cognition. Monographies*, 38 (1–2), 27–46.

Koržinek, Danijel 2020. Automating word segmentation. PINC Project. (https://pincproject2020.wordpress.com/2020/04/08/automating-word-segmentation/) (Accessed 2023-11-30).

Koržinek, Danijel & Chmiel, Agnieszka. 2021. Interpreter identification in the Polish Interpreting Corpus. *Revista Tradumàtica. Tecnologies de la Traducció*, 19, 276-288.

Kotze, Haidee. 2020. Converging What and How to Find Out Why: An Outlook on Empirical Translation Studies. In Vandervoorde, L. J. Daems & B. Defrancq (eds.). *New Empirical Perspectives on Translation and Interpreting*, New York: Routledge.

Kurz, Ingrid. 1993. Conference Interpretation: Expectations of Different User Groups. *The Interpreters' Newsletter*, 5, 13-21.

Laviosa, Sara. 1998. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, 43 (4), 557–70.

Lee, Tae-Hyung 1999. Speech proportion and accuracy in simultaneous interpretation from English into Korean. *Meta*, 44 (2), 260–67.

Liang, Junying, Yuanyuan Fang, Qianxi Lv & Haitao Liu. 2017. Dependency Distance Differences across Interpreting Types: Implications for Cognitive Demand. *Frontiers in Psychology*, 8.

Lin, Yumeng, Qianxi Lv & Junying Liang. 2018. Predicting Fluency with Language Proficiency, Working Memory, and Directionality in Simultaneous Interpreting. *Frontiers in Psychology*, 9.

Lim, Hyang-Ok. 2005. Working into the B Language: The Condoned Taboo? *Meta*, 50(4).

Mangiafico, Salvatore. 2023. rcompanion: Functions to Support Extension Education Program Evaluation. R Foundation for Statistical Computing, Vienna, Austria.

Mead, Peter. 2005. Directionality and Fluency: An Experimental Study of Pausing in Consecutive Interpretation into English and Italian. *Communication and Cognition. Monographies*, 38 (1–2), 127–46.

Monti, Cristina, Claudio Bendazzoli, Annalisa Sandrelli & Mariachiara Russo. 2005. Studying Directionality in Simultaneous Interpreting through an Electronic Corpus: EPIC (European Parliament Interpreting Corpus). *Meta*, 50 (4).

Ouyang, Lingwei, Qianxi Lv & Junying Liang. 2021. Coh-Metrix Model-Based Automatic Assessment of Interpreting Quality. In J. Chen & C. Han (eds.). Testing and Assessment of Interpreting: Recent Developments in China, New Frontiers in Translation Studies, Singapore: Springer, 179–200.

Öztürk, Asiye. 2020. The effect of directionality on performance and strategy use in simultaneous interpreting: A case of English-Turkish language pair. *RumeliDE Journal of Language and Literature Studies*, (18), 639–65.

Pavlović, Nataša. 2007. Directionality in Translation and Interpreting Practice: Report on a questionnaire survey in Croatia. *FORUM*, 5 (2), 79–99.

Piccaluga, Myriam; Nespoulous, Jean-Luc & Harmegnies, Bernard. 2005. Disfluencies as a window on cognitive processing. an analysis of silent pauses in simultaneous interpreting. *Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop*.10–12 September 2005, Aix-en-Provence, France, 151–155.

Plevoets, Koen & Defrancq, Bart. 2016. The effect of informational load on disfluencies in interpreting: A corpus-based regression analysis. *Translation and Interpreting Studies* 11(2), 202–224.

Pöchhacker, Franz. 2003. *Introducing Interpreting Studies*, London: Routledge.

Pöchhacker, Franz. 2012. Interpreting quality: Global professional standards. In W. Ren (ed.). Interpreting in the Age of Globalization: Proceedings of the 8th National Conference and International Forum on Interpreting, Bejing: Foreign Language Teaching and Research Press, 305–318.

Pradas Macías, Macarena. 2006. Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency. *Interpreting*, 8 (1), 25–43.

R core team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rennert, Sylvi. 2010. The impact of fluency on the subjective assessment of interpreting quality. *The Interpreters' Newsletter,* 15 (2010), pp. 101-115.

Russo, Mariachiara; Bendazzoli, Claudio & Sandrelli Annalisa. 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), 221–54.

Seleskovitch, Danica & Marianne Lederer. 1989. *Pédagogie raisonnée de l'interprétation,* Paris: Didier erudition.

Setton, Robin. 1999. *Simultaneous Interpretation*, Amsterdam/New York: John Benjamins Publishing Company.

Strömqvist, Sven, Victoria Johansson, Sarah Kriz, Hrafnhildur Ragnarsdóttir, Ravid Aisenman & Dorit Ravid. 2002. Toward a cross-linguistic comparison of lexical quanta in speech and writing. *Written Language & Literacy*, 5 (1), 45–67.

Tang, Fang. 2018. *Explicitation in Consecutive Interpreting*. Amsterdam:John Benjamins.

Tóth, Andrea 2013. The study of pauses and hesitations in conference interpreters' target language output. Unpublished doctoral dissertation, Eötvös Loránd University, Budapest.
Venables, Bill & Brian D. Ripley. 2002. *Modern Applied Statistics with S, Statistics and Computing*. New York, NY: Springer.
Wang, Binhua & Tao Li. 2015. An empirical study of pauses in Chinese-English simultaneous interpreting. *Perspectives*, 23 (1), 124–42.
Whyatt, Bogusława & Nataša Pavlović. 2021. Translating languages of low diffusion: current and future avenues. *The Interpreter and Translator Trainer*, 15 (2), 141–53.
Wijffels, Jan, Milan Straka & Jana Straková. 2019. UDPipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit."
Yu, Wenting & Vincent J. van Heuven. 2017. Predicting judged fluency of consecutive interpreting from acoustic measures: Potential for automatic assessment and pedagogic implications. *Interpreting*, 19 (1), 47–68.
Zwischenberger, Cornelia. 2010. Quality criteria in simultaneous interpreting: an international vs. a national view. *The Interpreters' Newsletter*, 15 (2010), 127-142.

Marta Kajzer-Wietrzny
Adam Mickiewicz University
Faculty of English
ul. Gruwaldzka 6
60-780 Poznań, Poland
e-mail: kajzer@amu.edu.pl

Agnieszka Chmiel
Adam Mickiewicz University
Faculty of English
ul. Gruwaldzka 6
60-780 Poznań, Poland
e-mail: achmiel@amu.edu.pl