Appeared in *Perspectives* published online 5 Feb 2020
https://doi.org/10.1080/0907676X.2020.1718170

Variation in non-fluencies in a corpus of simultaneous interpreting vs. non-interpreted English

Daria Dayter (University of Basel)

daria.dayter@unibas.ch, ORCID 0000-0001-5054-5330

The present study aims to investigate the occurrence of various non-fluencies (mispronounced words and hesitations; self-repairs and editing terms; silent pauses; repetitions; fillers) in interpreted and non-interpreted, spontaneously produced English. The material for the study is the English component of a parallel bidirectional corpus of Russian-English interpreting of political discourse, consisting of approx. 130,000 words from 77 speech events. The instances of non-fluencies have been automatically extracted from the corpus, with the exception of self-repair, which was subject to manual annotation. The figures for the two subcorpora were compared using the Mann-Whitney U test, chi-square test, and Fisher's exact test, as appropriate. The results show that (1) interpreted English has more disfluencies overall, and serial truncations specifically; (2) the number of repaired disfluencies is lower in interpreted English; (3) and interpreted English has fewer fillers and disfluent repetitions than non-interpreted English. The results on editing terms are inconclusive. While the first finding conforms to the predictions in the literature on SI, the other two can be ascribed to differences in style among interpreters.

Keywords: simultaneous interpreting; corpus studies; Russian-English; disfluency; pauses; repair

Acknowledgements: This study has been conducted with the support of a grant from Freiwillige Akademische Gesellschaft Basel for research stay at the Aston University in spring 2019.

Appeared in *Perspectives* published online 5 Feb 2020
https://doi.org/10.1080/0907676X.2020.1718170

1. Introduction

Various speech perturbations, as Schegloff et al. (1977) called them – hitches, cut-offs, sound stretches, filler noises – are a natural part of spontaneous speech. They have attracted attention of cognitive scholars and linguists alike, and have been studied as a window into thinking processes as well as a reflection of the complex social organisation of conversation. The same features occur in the speech of consecutive and simultaneous interpreters. It is the frequency and co-occurrence of these non-fluencies in simultaneous interpreting (SI) that is the subject of the present paper.

I am using a parallel bidirectional corpus of Russian-English interpreting to study disfluencies, self-corrections, silent pauses, and filler noises in interpreted and non-interpreted English. The rationale for the study is the idea that translated or interpreted language is different to originally produced one on a number of dimensions (so-called translation, or interpreting, 'universals', Baker, 1993; Shlesinger, 2008). That one of these dimensions could express itself in non-fluencies is indicated by the existing comparative research on SI. I carry out a large-scale systematic study of non-fluencies in a 130,000-word corpus to find out whether there is any difference between their frequency and types in interpreted vs. non-interpreted English, and what possible explanations could be offered for the findings.

The term 'non-fluencies' describes the totality of speech disruption phenomena. In it, I include filled and unfilled pauses, self-repair, repetitions, and disfluencies. Disfluencies are defined as "phenomena that interrupt the flow of speech and do not add propositional content to an utterance" (Gósy, 2007, p. 93). This distinction is important because, for example, appropriateness repairs may be delivered fluently, but nevertheless constitute self-correction and fall within the scope of this study. Similarly, a disfluency such as a truncated word or a false start may be accompanied by a repair, but need not be.

I address the following research questions:

1) What non-fluencies are present in the corpus?
2) Are disfluencies followed by self-repair or not, and can a pattern be observed?
3) Is there a difference between the original English and the interpreted English subcorpora regarding the findings of RQ 1 and 2?

The article comprises four sections. The first section provides an overview of the existing literature on non-fluencies in language as a whole, as well as the chosen model of output monitoring. The second section focuses on the relevant literature specifically in simultaneous interpreting. It is followed by the analysis of non-fluencies in the English component of the bidirectional corpus of

https://doi.org/10.1080/0907676X.2020.1718170

Russian-English interpreting, that is, the English source texts (ST) and the English target texts (TT[1]). The fourth section offers a discussion of the results in view of pragmatic and cognitivist research, and proposes an explanation of the emerging pattern.

2. Output monitoring and repair in conversation

In the following, I review two different approaches to the study of non-fluencies. One takes a cognitivist point of view and engages with speech monitoring models, attempting to discern speech production mechanisms on the basis of their temporary hitches. The other stems from the Conversation Analysis tradition and is interested mostly in self-repair as evidence of cooperation among interlocutors.

In the cognitivist tradition, on-line editing of speech errors is taken as evidence for the existence of an output monitor of speech, either on the level of speech production, or as inner speech. One of the most fruitful models has been put forward by Levelt (1983). Levelt works with a selection of 959 repairs from a corpus of audio recordings of an elicitation task in Dutch. He weighs his data against Nooteboom's (1980) Main Interruption Rule, which states that a monitoring speaker stops the flow of speech immediately upon detecting the occasion of repair (i.e. with minimum latency). Levelt's conclusion is that repair indeed occurs immediately upon the detection of trouble. This and other studies (Blackmer & Mitton, 1991) support the perceptual loop theory of monitoring – meaning the speaker parses his or her inner speech rather than waiting for a finished utterance. The perceptual loop theory has important implications for simultaneous interpreting, which places additional time constraints on any correction process.

To describe the foci of monitoring, Levelt (1983, pp. 51-55) developed a classification of repairs in his extensive data. He outlined four main categories, based on the nature of repair content and placement (the examples marked [p] are taken from the paper on SI by Petite, 2004):

- In D-repairs, the speaker decides to express a different message or the same message in a different order, interrupts him- or herself and starts anew. They accounted for 1% of repairs.

  [p] "und er hat sich im besonderem mit e::: Ge-General e::: er ist er ist ein Schw- ein Schwede von e::: Herkunft";

[1] Although it is a corpus of interpreting, hereafter I will refer to the transcriptions of speeches as "texts", to enable further use of traditional terminology such as "target/source text".

- In A-repairs, the speaker corrects for appropriateness (not in the pragmatic sense, but in regard to how well the chosen reference fits into the larger discourse). They accounted for 30% of repairs. These are subdivided into
    - AA-repairs, to correct the ambiguity of the message

      "We start in the middle with… in the middle of the paper with";
    - AL-repairs, to shift the level of terms, e.g. from less to more precise

      [p] "Ich hoffe ebenso dass Ihre Tag Ihre Zeit die Sie in Dublin verbringen";
    - AC-repairs, to correct coherence

      [p] "Now what is the third element of the family and the body the the soul and the body. The body e::: is the economy of the family";
    - ALC-repairs, used to describe fuzzy cases between AL and AC repair;
- In E-repairs, errors are dealt with. They accounted for 41% of repairs. These repairs are subdivided into
    - EL-repairs, to correct a lexical item

      [p] "Exekutivsekretär des Wirtschafts-und Sozialrats der Vereinten Nationen für Europa, Entschuldigen Sie der Wirtschaftskommission für Europa";
    - ES-repairs, to correct sentence structure

      [p] "That means an unconditional e::: m::: acceptance that does not determine is not determined be::: or is not limited by performance";
    - EP-repairs, to correct a mispronunciation or a slip of tongue

      "A unut, unit from the yellow dot";
- C-repairs, or covert repairs, involve repetitions or editing terms when it is impossible to tell what level of language was being monitored. They accounted for 25% of repairs.

  "Then right , uh grey";
- Finally, the additional R-repair is a bin category for unclassified cases; 3% of Levelt's examples.

This simple yet all-encompassing taxonomy has received much uptake in self-repair research and has also been applied in SI studies (see next subsection). However, Levelt's (1983) interest on repairing disfluencies makes the reason for repair the focus of investigation, leaving aside the surface-level features of the trouble moments. The surface form of repair is the focus of the work by Gósy (2007) and Magno Caldognetto et al. (1982, ctd. in Tissi, 1999). They propose a distinction between disfluencies stemming from uncertainty vs. error-type disfluencies, or silent pauses vs. non-fluencies.

https://doi.org/10.1080/0907676X.2020.1718170

Linguists have also studied non-fluency and repair in the context of Conversation Analysis (CA). Conversation analysts demonstrated that repair is far from an emergency communication breakdown but is instead "an organised set of practices through which participants in conversation are able to address and potentially resolve problems of speaking, hearing or understanding" (Sidnell 2010, p. 110).

The question of locating the repairable is not a trivial one for a corpus study of non-fluencies since it is not always overt. Indeed, the very term 'repair' is misleading, because it is "sometimes found where there is no hearable error, mistake, or fault" (Schegloff et al., 1977, p. 363). To cover as many instances of speech trouble as possible, I have manually annotated the corpus transcripts for self-repair, and semi-automatically for disfluencies.

The findings of the CA branch of research concern the functional aspect of non-fluencies. Kaur (2011), in a study of naturally occurring conversation among speakers of English as a foreign language, established that self-repair can work as an explicitation strategy, improving clarity and comprehensibility of speech. The global strategy of explicitation has been shown to be extremely common in SI (Riccardi, 2005; Dayter forthcoming), which could perhaps be accounted for by self-repair. Fox and Jasperson (2014) demonstrate how repair, embedded within the repetition of a large syntactic component, could be a strategy for delaying the end of the turn. This has interesting implications for the study of SI where the interpreter has no need to compete for the floor and therefore might exhibit fewer repetitions. Finally, silence has a variety of functions in talk as well: it can preface a disaffiliation (Goodwin & Heritage, 1990), be a resource for exiting a conversational sequence (Schegloff, 2009), or prompt self-repair (Nakane, 2011).

The reviewed findings feed into the Hypothesis A formulated below, namely, that the corpus of SI is expected to have more non-fluencies on the whole than the original English. The next section provides an overview of non-fluency and repair research in simultaneous interpreting within the two paradigms.

3. Non-fluencies and repair in simultaneous interpreting

Both of the research traditions above have been brought to bear on SI, although the cognitivist approach appears more popular among interpreting scholars. The famous Effort Model (Gile, 1995) and, more generally, the consideration of cognitive strain, provide a ready theoretical basis for non-fluencies in interpreters' speech. Bakti (2008), for instance, draws a parallel between SI and speech production in noisy conditions, which had been shown by psycholinguistic research to result in non-fluencies. Various comparative studies confirm that disfluencies can be traced back to cognitive load and cognitive skills (Bakti, 2008; Cecot, 1996; Collard & Defrancq, 2019; Tóth, 2011).

Given that interpreters find themselves under a heavier cognitive load than speakers who speak spontaneously, several studies have tested the hypothesis that SI contains more non-fluencies than non-interpreted speech, in line with Hypothesis A below. Pöchhacker (1995), working with the speeches from a three-day conference interpreted bidirectionally in English-German, observes more slips and shifts in the speech of interpreters, while speakers had more uncorrected slips. Collard and Defrancq (2019) see the same trend for more disruptions in interpreters' speech in regard to filled pauses, lengthenings, false starts, and longer silent pauses. Tissi (1999), on the material of German-Italian, finds that certain categories – false starts, vowel and consonant lengthenings – are more common in the target texts. These studies inform the Hypothesis C formulated below concerning the greater amount of filled pauses and repetitions in interpreters' speech. Tissi remarks, however, that there is large variation among individual interpreters when it comes to other categories, such as filled pauses and restructuring. This led Van Besien & Meuleman (2014) to suggest that interpreters might choose to adopt different individual styles: 'lean' or 'abundant', respectively associated with more fluent delivery that avoids corrections, or with the generous use of explicitation, appropriateness repairs, etc.

Levelt's (1983) classification has been used and adapted by Petite (2004, 2005) in a pragmatics-oriented study of repair. To adequately describe the repairs at hand, Petite (2005) introduced two additional repair categories. One is mid-articulatory (MA) repair, that is, repair that starts after the word articulation began (in contrast to Levelt's post-articulatory/pre-articulatory dichotomy). The other is the grammatical error repair (EG) that she apparently uses instead of syntactic error (ES).

Petite's (2004) findings for SI are proportionally similar to those by Levelt (1983): around 28% of post-articulatory repairs fall into the A (appropriateness) category, 32% into E (error), 11% into D (different message), and 29% into covert/mid-articulatory. The similarity might be due to the fact – apart from the obvious observation that SI and spontaneous speech function identically in terms of monitoring – that SI in Petite's corpus was of the 'with text' variety. As she specifies (Petite 2004:106), various materials had been made available to the interpreters prior to their work sessions. In my description of post- and mid-articulatory self-repairs, I will use the same classification to test its applicability to free SI, which conceivably might differ.

The study that most closely resembles the present one is the description of non-fluencies in the EPIC corpus by Bendazzoli et al. (2011), so I will discuss it in more detail. The crucial difference between the two lies in the choice of language pair(s) and in the approach to identifying the non-fluencies. First of all, Bendazzoli et al. (2011) work with six subcorpora of SI in English, Italian, and Spanish. It has been shown that not only directionality affects linguistic choices, but also that repair is realised differently in different languages, even spoken by the same person (Rieger, 2003). Therefore, an addition of a new source language and interpreters with a new linguistic background is a valuable contribution to the body of knowledge. Moreover, with the exception of Hungarian, the reviewed literature draws on closely related Germanic and Romance languages. Russian-English interpreting might yield different findings. Secondly, Bendazzoli et al. (2011) take into account only such ‘disfluencies’ (the authors’ preferred term) and repairs that have searchable triggers (mispronunciations and truncated words), and likely leave out corrections that do not follow a hearable error.

EPIC is a corpus of European Parliament speeches consisting of six subcorpora, approx. 178,000 tokens. Passages were automatically extracted from this corpus to investigate the nature of disfluency (single or serial), the presence of editing terms (verbal material added between mispronounced or truncated words and their repair, e.g. “I meant”), and the point where the articulation of truncated words stopped (Bendazzoli et al., 2011, p. 284). On the basis of this data, the authors also tested the hypothesis about the prevalence of disfluency in SI over non-interpreted language.

The results are ambiguous. Indeed, SI has more disfluencies than the source language (SL) text, but not without exception. English source texts have more truncated words, and fewer pronunciation troubles, than Spanish or Italian target texts. SI also has more disfluencies than comparable speeches delivered in the same language. Mispronunciations were generally not corrected in SI, but truncated words were completed. Truncated words most commonly resulted from interpreters trying to approximate the desired lexical item, and there was a pronounced language family effect for this type of disfluency (Spanish speakers favouring editing terms). The general conclusion is that production problems in original and interpreted speeches are of a different kind: original speakers incur problems related to phonological effects, and interpreters are more likely to incur syntactic and lexical planning errors (Bendazzoli et al., 2011, p. 297). The findings of this study inform the Hypothesis B below.

On the basis of the literature review above, I formulated the following hypotheses:

A. The SI subcorpus will exhibit more non-fluencies on the whole than the original subcorpus.

B. The SI subcorpus will have fewer repaired disfluencies and fewer editing terms than the original subcorpus.

C. The SI subcorpus will have fewer filled pauses and repetitions than the original subcorpus.

4. Non-fluencies in the Russian-English corpus

4.1 Data and methodology

The corpus used for this study was compiled from material of the United Nations Web TV (http://webtv.un.org/) as well as broadcasts of the video news agency RT. WebTV airs various UN events such as General Assembly, press conferences, press briefings etc. with the original soundtrack and also tracks in all the official UN languages, including Russian and English. For the corpus, I used only the events originally held in either Russian or English.

The RT YouTube channel yielded a collection of press conferences, briefings, and interviews by Russian, American and British politicians and public figures that were broadcast with dual track, SI and original language. Since these are televised events interpreted off-site during the broadcast, this component of the corpus falls under the category of live television interpreting. It involves an interpreter working in conditions similar to a traditional interpreter, i.e. wearing headphones and speaking into a microphone, and under the same time pressure due to the live broadcast situation. For that reason, both components of the corpus – true live SI and live television SI – are treated as homogenous in this study (for a detailed discussion, see Dayter 2018). At the moment, both interpreting directions consist of approx. 60% free SI (16 out of 28 texts, 63% of total tokens in En>Ru vs. 29 of 49 texts, 60% of total tokens in Ru>En).

The corpus consists of two Russian components, a non-interpreted and an interpreted one (hereafter RuOr and RuSI), and two English components, a non-interpreted and an interpreted one (hereafter EnOr and EnSI). For this study, I use the EnSI and EnOr subcorpora consisting of 77 individual speech events (see Figure 1 for token counts). Speeches are divided into separate files by interpreter, i.e. there is no change of interpreter within one transcript.

The speeches have been transcribed using a simple orthographic transcription enriched with the features of spoken speech relevant to the study of non-fluencies (see Table 1). Pauses are annotated in half-second increments, with the shorter pauses annotated only if the transcribers have judged them as marked rather than unmarked features of speech flow. Descriptions of non-verbal events on tape, such as 'loud bang', allow me to see if a non-fluency has an external, situational cause, as opposed to being a processing trace.

Figure 1. Composition of the corpus[2]

Table 1. Non-fluency notation in the corpus

Notation	Meaning	Example
<.>, <1>, <2.5>, <5> etc.	Silent pause (number indicates length in seconds, rounded up to 0.5; <.> indicates a micropause shorter than 0.5 sec)	*let's get <.> down to the original agenda* *namely the situation in Syria <3> we start <2> our briefings*
<rep>	Tag used to mark the repetition of a word as a disfluency (not for emphasis)	*within the confines of of <rep> the dinner*
<dsf>	Tag following disfluencies that do not add propositional content to the utterance: mispronounced words, truncated words, elongations, false starts	*I will al-also <dsf> preside over the high-level meeting in December*
<scr>	Tag following an instance of self-repair, usually occurs following a disfluency	*and our fountry <dsf> fellow countrymen <scr> in the South-East*
-	Elements of a truncated word or false start	*responsibilities of nations and the resp- <dsf> rights <scr> of individuals*
<loud bang>	Description of nonverbal on-tape events	*but obviously the nature <loud noise> of a sort of security issue*
<unclear>, <please>	Placeholder for cases when the audio source is (partly) incomprehensible	*this is a serious eh fighting eh it adds to eh already important <unclear> international law*

To prepare the data for the analysis, all instances of self-repair, disfluencies, repetitions, silent pauses, fillers, and truncated words were extracted from the two English subcorpora, retaining information about their immediate co-text, interpreted or non-interpreted status, and the individual texts. The self-repair examples were then manually annotated in accordance with the repair

[2] The size of the corpus is slightly different from the one reported in an earlier study, Dayter (2018), because the composition has been adjusted to balance free/with text SI.

taxonomy by Levelt (1983) as modified for SI by Petite (2005) (see section 2 above)[3]. I also manually annotated all non-fluency examples for the presence or absence of editing terms. To obtain a count of uncorrected disfluencies, I excluded all hits where the tag <scr> occurred within 50 characters to the right of the tag <dsf>.

Since the transcription of pause length in the corpus is precise to 0.5 seconds, and to ensure that the pauses are disfluent and not a feature of the individual speech rate, a very high threshold of L > 2 seconds was set for pause extraction. This is considerably longer than in cognitive studies of silent pauses in SI which relied on more granular transcription (e.g. 0.2 sec in Tissi, 2000; Collard & Defrancq, 2019), and is intended to minimise false positives.

To quantify each feature, the raw counts of all non-fluencies were normalised. The normalisation basis for self-repair, disfluencies, repetitions, and the truncation elements was the total number of tokens per text. The normalisation basis for silent pauses (measured in seconds) was the length of individual recordings in seconds.

The comparison of the resulting data (comparable lists of relative feature counts per text for disfluencies, repetitions, fillers, silent pauses) was carried out using the Mann-Whitney U test to obtain the p values, and the Vargha and Delaney A to compute the effect size. The self-repair categories per Levelt (1983) were compared pairwise per category using the Chi-square test with Yates's correction and the phi measure of effect size. The amount of corrected vs. uncorrected disfluencies in the two corpora was compared using Fisher's exact test, as one of the cell values was quite low (N=5). The threshold for significance was set at $p<0.05$. All statistical analysis and plot generation were done using R (R Core Team, 2016; Fan, 2017; Torchiano, 2018).

4.2 Quantitative analysis of disfluencies

The total number of disfluencies, i.e. detectable errors of speech, is significantly higher in the EnSI subcorpus than in the EnOr, confirming the original hypothesis ($p<0.001$, $A=0.19$ (large effect)). The information on frequency and distribution of disfluencies can be found in Table 2 and Figure 2.

Table 2. Relative frequencies of disfluencies in the EnOr and EnSI subcorpora

Subcorpus	Relative frequency of dsf, per subcorpus (%)	Relative frequency of dsf, per text (%)	
		M	SD
EnOr	0.34	0.38	0.37

[3] I am grateful to José Belém who acted as the second coder.

https://doi.org/10.1080/0907676X.2020.1718170

EnSI	0.84	0.9	0.6

Figure 2.

The category of disfluencies is comprised of vowel elongations and other hesitations, false starts, and mispronounced and truncated words, as one can see in the following examples:

1) that there's no common vision between the West a-and Russia on Ukraine that international mediators are not needed
2) further down in the transcript the second time that he men- mentions it he says that Prime Minister Netanyahu himself asked me to try
3) introductory memorandum regarding the strength of enc- of a sort of <unclear> there are in fact today
4) financial <.> contributions that are necessary for an implemesntation

Silent pauses as such were not tagged as disfluencies unless they were accompanied by other troubles, as in example 5. For that reason, the analysis of silent pauses is presented separately in section 4.5.

5) that you believed was not fair aaah s-sssss <2> so I-I impose sanctions

It is interesting that although the number of various truncations is significantly higher in interpreted English, if we look at serial vs. single truncations, the two corpora are more similar (Table 3, Figure 3). This is done by extracting every truncated element instead of treating them as a single disfluency: thus, example 6, which would be counted as N_{dsf}=1 in the analysis above, in this case is $N_{hesitation}$=2.

6) let us apply legal pre-pre- eh ehm eh prescriptions

Table 3. Relative frequencies of hesitations in the EnOr and EnSI subcorpora

Subcorpus	Relative frequency of hesitation, per subcorpus (%)	Relative frequency of hesitation, per text (%)	
		M	SD
EnOr	0.84	1.1	1.12
EnSI	1.04	1.1	0.6

Figure 3.

https://doi.org/10.1080/0907676X.2020.1718170

There is no significant difference found in the two subcorpora (p>0.05, A=0.34 (small effect)). This is in line with the findings of earlier research, which predicts that interpreters will produce disfluencies which indicate difficulties with lexical and syntactic planning, but not disfluencies whose function is to hold the floor. Long instances of word-initial hesitations, represented in the corpus by multiple truncated fragments, are examples of the latter category. The same is true for filler noises which are the subject of section 4.4.

4.3 Quantitative analysis of self-repair

This section will look at self-repair in the corpus in two steps. First, I will consider all instances of self-repair, including the ones that do not follow a hearable error and were identified by manual coding. Second, I will compare the rates of correction of the disfluencies, that is, only consider those self-repairs that follow hearable errors. I will also talk about editing terms.

The total number of self-repairs per text is higher in EnSI than in EnOr, in line with the hypothesis that interpreted texts will exhibit more non-fluencies (p<0.01, A=0.26 (large effect)). Below is an example of a self-repair that does not follow a hearable disfluency and would therefore have escaped notice in automatic data extraction:

7) I would ask all the participants to speak quite slowly, particularly when asking questions or expecting a reaction, because <u>of a time a possible time-lapse</u> of video-link

All self-repair in the English subcorpus was categorised according to the Petite-Levelt taxonomy of repair, distinguishing among four types of appropriateness repairs, four types of error repairs, message changes, and mid-articulatory repairs (which are by definition always lexical). The results are reported below in Table 4.

Table 4. Relative frequencies of repair types in the EnOr and EnSI subcorpora

Category		Repair frequencies in English (%)				Chi-square
		EnOr		EnSI		
		Per subcategory	Total per category	Per subcategory	Total per category	
A	AA	2.9		5.5		
	ALC	36.2		18.8		
			39.1		**24.4**	p<0.05*, phi=0.12 (small)

E	EG	10.1		25.8		
	EL	17.4		23.2		
	EP	7.2		1.8		
	ES	8.7		0		
			43.5		**50.9**	p>0.05
D		4.3	**4.3**	6.6	**6.6**	p>0.05
MA		13	**13**	18.1	**18.1**	p>0.05
Total		100%	100%	100%	100%	

When the corpora are compared pairwise in regard to general repair categories (A, E, D, and MA), the only significant difference is found for the appropriateness errors. The subcategories of A-repairs are distributed similarly, with coherence and lexical appropriateness repairs being the most common in both, although in EnOr ALC-repairs account for a larger proportion of A-repairs (92% of all A-repairs in EnOr and 77% of all A-repairs in EnSI). The prevalence of ALC in EnOr can be expected: such repair has to do with the generation of the message, which is the responsibility of the speaker. The interpreter, acting as an animator, is responsible for lexical and syntactic choice in the target language (that would be handled in E-repairs), but does not have the authority to adjust the appropriateness.

Not every self-repair accompanies a hearable error and, conversely, not every disfluency has been repaired. The comparison of the rates of correction of disfluencies in the corpora shows that, although speakers repair overt disfluencies slightly more often than interpreters, the difference between interpreted and non-interpreted language is not significant (see Table 5; $p>0.05$). The lower rate of correction in TT compared to ST can be due to the time pressure and cognitive strain, which prevent interpreters from introducing additional language material or, indeed, from closely monitoring the output.

Table 5. Repaired and unrepaired disfluencies in the EnOr and EnSI subcorpora

Subcorpus	Repaired dsf (%)	Unrepaired dsf (%)	
EnOr	95	5	100%
EnSI	94	6	100%

Finally, very few editing terms appear in either corpus. Only two instances of repair in EnOr and one in EnSI were accompanied by additional language material, in all cases some variation on 'sorry':

8) recently set up a committee on eradic- I'm sorry, the UN committee on education of all forms

This is considerably lower than what Bendazzoli et al. (2011) found in their EPIC material, where in the English subcorpora between 22% and 37% of truncated words were accompanied by an editing term. The number was a lot lower for mispronounced words, between zero and 5%, which still places EPIC above the En-Ru corpus. However, Bendazzoli et al. (2011) note that English had the lowest editing term frequency of all language corpora, with Romance languages showing a particular affinity to acknowledging the trouble. This dispreference of English appears to have been amplified in my data.

All in all, the second hypothesis – that interpreted English would have fewer editing terms and repaired disfluencies than non-interpreted English – could not be confirmed.

4.4 Filler noises and repetitions

Filler noises – such as *eh, uh, erm* – occupy an uncertain position in the existing classifications of non-fluencies. Cecot (1996) considers them a subtype of hesitation pauses; Tissi (2000), on the contrary, claims that vocalized hesitations are a subtype of filled pauses; in Gósy's (2007) taxonomy, hesitations and fillers are on the same hierarchical level. Levelt (1983) even classed fillers as indicative of a covert repair.

In the present study, I do not include fillers as a subtype of self-repair since a covert repair is impossible to identify in product-oriented study. Instead, fillers are considered as a separate category of variation between interpreted and non-interpreted language. In line with the Conversation Analysis view, the expectation is that speakers produce more fillers than interpreters because their primary function is to hold the floor and indicate that the conversational turn is not over. Since an interpreter does not have the authority to manage the floor, these devices will occur less frequently. Repetitions, which in this study have been defined as non-emphatic, exclusively disfluent repetitions, fulfil the same function and are also expected to occur less frequently in EnSI.

The findings confirm this hypothesis. Fillers are significantly more common in EnOr ($p<0.01$, A=0.7 (medium effect)), and so are repetitions ($p<0.01$, A=0.7 (medium effect)). However, individual texts vary greatly in the number of fillers: M=3.5%, SD=3 for EnOr and M=1.7%, SD=1.7 for EnSI.

Table 6. Relative frequency of fillers and repetitions in the EnOr and EnSI subcorpora

Subcorpus	Relative frequency of fillers (%)	Relative frequency of repetitions (%)

EnOr	2.5	0.2
EnSI	1.2	0.1

4.5 Silent pauses

Finally, unfilled pauses are also treated as a separate kind of non-fluency. It is difficult to offer a hypothesis concerning silence since it is potentially extremely multifaceted and lacks overt indicators of the underlying causes. Grosjean (1980, ctd. in Cecot, 1996, p. 66) stated:

> There may be 40 or 50 different variables that can create a silence in speech. A silence may mark the end of a sentence, you can use it to breathe, you can use it to hesitate: there may be 10 or 15 different things happening during silence.

Conversation analysts interpret pauses strategically – for example, in interviews a pause can serve as a trigger of other-initiated self-repair (Nakane, 2011). For simultaneous interpreting, it has been suggested that silent pauses, especially long ones, of the kind under discussion here, are a sign of high attention to input and a temporary withdrawal of attention from the formulation task (Setton, 1999, p. 247). These suggest competing interpretations. The CA approach leads us to expect fewer silent pauses in EnSI, while the cognitivist approach predicts silent pauses in free EnSI where all input is auditory and cognitive strain is pronounced.

When the summative length of all disfluent pauses (L>1.5 sec) is compared across EnOr and EnSI speeches, there are no significant differences between interpreted and non-interpreted language (see Table 7). The distribution of silent pauses is extremely varied, with several interpreted speeches involving disfluent pauses of up to 10 seconds – these are the outliers visible in Figure 4. A closer inspection of the transcription has shown that the interpreter omitted large portions of content in these texts and generally appeared to be struggling to maintain synchronicity.

Table 7. Silent pauses in the EnOr and EnSI subcorpora

	Length of disfluent silences per subcorpus, %	SD
EnOr	4.4	3.8
EnSI	6.5	6.8

Figure 4.

Appeared in *Perspectives* published online 5 Feb 2020
https://doi.org/10.1080/0907676X.2020.1718170

5. Discussion: non-fluencies in simultaneous interpreting

To sum up, a comparison of various non-fluencies in interpreted and non-interpreted English has yielded the picture displayed in Table 8. Plus symbols indicate a prevalence of the feature in a subcorpus, minus symbols indicate that no significant difference was found. In the case of editing terms, too few occurrences were found in the corpus to draw any comparative conclusions.

Table 8. Summary of the non-fluencies comparison of the EnOr and EnSI subcorpora

	Non-fluencies						
	Disfluencies	Repaired disfluencies	Editing terms	Silent pauses	Serial hesitations	Repetitions	Fillers
EnOr		-	-	-		+	+
EnSI	+	-	-	-	+		

Hypothesis A stating that interpreted English will exhibit more non-fluencies on the whole has been confirmed.

Hypothesis B was confirmed in regard to the lower number of repaired disfluencies in EnSI, and the results were inconclusive regarding editing terms.

Hypothesis C stating that interpreted English would have fewer fillers and disfluent repetitions than non-interpreted English has also been confirmed.

Despite the fact that the general trends predicted by the literature hold in the En-Ru corpus, the results on a more specific level diverge in several places. Percentages per repair category, when compared to the findings of Levelt (1983) and Petite (2004), demonstrate the same ranking but different weighing of categories (Table 9). I found fewer A-repairs and more E-repairs than both other authors.

Table 9. Overview of research results of self-repair categories

Type of repair	Levelt (1983), DutchOr	Petite (2004), EnSI	Present study, EnSI
Postarticulatory A	30%	27.7%	24.4%
Postarticulatory E	41%	31.9%	50.9%
Postarticulatory D	1%	11.7%	6.6%
Covert/mid-articulatory	25%	28.6%	18.1%

https://doi.org/10.1080/0907676X.2020.1718170

The contrast with Levelt's (1983) study can at least in part be ascribed to the different data (he studied spontaneously produced Dutch) and definition of category (covert and mid-articulatory repairs are not an entirely overlapping set and are very difficult to detect). The study by Petite (2004), however, is very similar to the present one with the exception of source languages. This difference in results, then, underscores once again the vagueness of categories and the variability in interpreters' performance. Van Besien & Meuleman (2014) described two distinct interpreter styles: lean and abundant. The increase in E-repairs in my corpus can be a reflection of interpreters' abundant styles, whereby they choose to repair errors that another interpreter might have left unrepaired for the sake of a more fluid delivery.

Some conclusions regarding the effect of directionality can be drawn from comparing the results of the present study (Ru-En) to the results of the disfluency analysis in EPIC (En-It-Sp, Bendazzoli et al., 2011). Although 'mispronounced words' were not quantified as a separate category here, 'truncated words' can be compared across the two studies and yield remarkably similar results (Table 10). It appears that Russian ST produces results more similar to Italian than Spanish ST in EPIC. With regard to editing terms, my results are also closer to Italian ST. Since no particular typological connection can be traced in this case, it is likely that the similarity is again due to interpreters' individual style.

Table 10. Overview of research results on truncated word frequency

Subcorpus	Truncated words (%)	
	EPIC	Present study
EnOr	0.756	0.86
EnSI	1.013 (It-En)	1.05 (Ru-En)
	0.884 (Sp-En)	

To sum up, non-fluencies in this corpus are considerably more similar to the results of other studies of interpreting variation than, for example, lexical complexity or POS-collocations (Dayter 2018; 2020). One could propose a tentative observation that non-fluencies, therefore, are less dependent on language directionality than lexis-oriented variables. This is meaningful for the choice of variables in practical applications such as translation status classification tasks, native language influence detection, or speaker profiling.

6. Conclusion and avenues for further research

Let us review the findings within the wider context of the theories of simultaneous interpreting. The cognitivist theories consider pauses and other non-fluencies as traces of cognitive activity. The general findings of this study therefore conform neatly to the expectation that the strain on the interpreter's processing capacity, which is greater than the strain occasioned by speaking normally, would cause increased non-fluency. It is substantiated by the variationist studies cited earlier, as well as research on the relationship between certain features of ST and the disfluency rate in TT (e.g. Plevoets & Defrancq, 2016 on delivery rate, lexical density, numbers, and sentence length). The parallel study of the Russian component of this corpus produced the same results with respect to total non-fluency prevalence in TT (Dayter under review).

However, it remains unclear why certain non-fluencies in the sample did not differ in SI and Or (silent pauses), or were higher in Or (repetitions and fillers). Above I have proposed a pragmatic solution – the reduced need to hold the floor in interpreted speech; and the 'mixed bag' solution – the functional decrease in pauses vs. the processing increase in pauses balance each other out. These ad-hoc explanations do not always conform to earlier findings, for example, on silent pauses which have been shown to be more common in SI by other authors.

A tentative explanation can be sought in the work on interpreting universals, although, of course, the study relying on only one language pair cannot offer strong indications in any direction. Shlesinger (1989) proposed an equalising universal to explain the systematic variation between non-interpreted and simultaneously interpreted language. She hypothesised that the process of interpreting exerts an equalising effect on oral and literate features of source texts, whereby oral texts receive more literate features and vice versa. Although non-fluencies on the whole are, of course, a feature of orality, the texts in the corpus are not completely homogenous in terms of their placement on the orality-literacy continuum. An analysis of non-fluencies in individual texts can show whether this is a dimension of variation that explains contradictory results, once the corpus has been extended to include a sufficient amount of material on each type of text (prepared written speeches, prepared speeches based on notes, oral reports, spontaneous speeches, prepared Q&A sessions, spontaneous Q&A sessions, to name just a few).

At the moment, however, the most convincing explanation remains the individual difference between interpreter styles. This is not to say that such styles are set in stone. It is more likely that each interpreter possesses a repertoire of styles ranging from lean to abundant and employs them depending on the context. Different situations exercise different requirements in regard to the performative aspect of SI as opposed to the content-centred aspect. In one situation, an interpreter may favour fluent performance at the expense of error repair, and in another an abundant style with generous explicitations and appropriateness repairs to accommodate an audience with a different background. A corpus-based investigation into such styles is a promising research direction that can be addressed with the help of corpora supplying rich metadata, e.g. an interpreting analogue of the TransBank (Ustaszewski & Stauder, 2017).

References

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.* Amsterdam: Benjamins, 233–250.

Bakti, M. (2008). Speech Disfluencies in Simultaneous Interpretation. In D. De Crom (Ed.), *Selected Papers of the CETRA Research Seminar in Translation Studies 2008*. Leuven, 1-17.

Bendazzoli, C., Sandrelli, A., & Russo, M. (2011). Disfluencies in simultaneous interpreting: a corpus-based analysis. In A. Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based Translation Studies: Research and Applications*. London: Continuum, 282-306.

Blackmer, E., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*, 173–194.

Cecot, M. (1996). Pauses in Simultaneous Interpretation: a Contrastive Analysis. *The Interpreter's Newsletter*, *11*, 63–85.

Collard, C., & Defrancq, B. (to appear 2019). Disfluencies in simultaneous interpreting; a corpus-based study with special reference to sex. In L. Vandevoorde, J. Daems & B. Defrancq (Eds.), *New Empirical Perspectives on Translation and Interpreting*. London: Routledge.

Dayter, Daria. (2018). Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN). *FORUM: International Journal of Interpretation and Translation, 16 (2)*, 241-264.

Dayter, Daria. (2020). Collocations in non-interpreted and simultaneously interpreted English: a corpus study. In J. Daems & B. Defrancq (Eds.). London: Routledge.

Dayter, Daria. (forthcoming). Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological complexity, and position in speech event. *Meta: Journal des traducteurs.*

Dayter, Daria. (under review). Pausing and hesitating in Russian: findings from a corpus simultaneous interpreting. *Vestnik RHGA*.

Fan, F. Y. (2017). powerAnalysis: Power Analysis in Experimental Design. R package version 0.2.1. https://CRAN.R-project.org/package=powerAnalysis

Fox, B. A., & Jasperson, R. (2014). A syntactic exploration of repair in English conversation. In P. Davies (Ed.), *Descriptive and Theoretical Modes in the Alternative Linguistics*. Amsterdam: Benjamins, 77-134.

Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam: Benjamins.

Goodwin, C., & Heritage, J. (1990). Conversation Analysis. *Annual Review of Anthropology*, *19*, 283–307.

Gósy, M. (2007). Disfluencies and Self-monitoring. *Govor*, *26*, 91–110.

Kaur, J. (2011). Raising explicitness through self-repair in English as a lingua franca. *Journal of Pragmatics*, *43*(11), 2704–2715.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104.

Nakane, I. (2011). The role of silence in interpreted police interviews. *Journal of Pragmatics*, *43*(9), 2317–2330.

Nooteboom, S. G. (1980). Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech. In V. Fromkin (Ed.), *Errors in linguistic performance*. New York: Academic Press.

Petite, C. (2004). *Repair mechanisms in simultaneous interpreting : a corpus-based analysis of interpreters' deployment of processing resources (English/French/German)*. *PhD dissertation, Heriot-Watt University.*

Petite, C. (2005). Evidence of repair mechanisms in simultaneous interpreting. *Interpreting*, *7*(1), 27–49.

Plevoets, K., & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting. *Translation and Interpreting Studies*, *11*(2), 202–224.

Pöchhacker, F. (1995). Slips and Shifts in Simultaneous Interpreting. In J. Tommola (Ed.), *Topics in Interpreting Research*. Turku: University of Turku, 73-90.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Riccardi, A. (2005). On the evolution of interpreting strategies in simultaneous interpreting. *Meta: Translators' Journal*, *50*(2), 753–767.

Rieger, C. L. (2003). Repetitions as self-repair strategies in English and German conversations. *Journal of Pragmatics*, *35*, 47–69.

Schegloff, E. A. (2009). A Practice for (Re-)Exiting a Sequence: And/But/So + Uh(m) + Silence. In *Language in Life, and a Life in Language: Jacob Mey, a Festschrift*. Leiden: Brill, 365-374.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, *53*(2), 361–382.

Setton, R. (1999). *Simultaneous interpretation: a cognitive-pragmatic analysis*. *Benjamins translation library*. Amsterdam: Benjamins.

Shlesinger, M. (1989). *Simultaneous Interpretation as a Factor in Effecting Shifts in the Position of Texts on the Oral-Literate Continuum*. *PhD dissertation, Tel Aviv University.*

Shlesinger, M. (2008). Towards a definition of Interpretese: An intermodal, corpus-based study. In G. Hansen, A. Chesterman, & H. Gerzymisch-Arbogast (Eds.), *Efforts and Models in Interpreting and Translation Research: A tribute to Daniel Gile*. Amsterdam: Benjamins, 237-253.

Sidnell, J. (2010). *Conversation analysis. An introduction.* Malden: Wiley-Blackwell.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. London & New York: Wiley-Blackwell.

Tissi, B. (1999). Silent Pauses and Disfluencies in Simultaneous. *The Interpreter's Newsletter*, *10*, 62–63.

Torchiano, M. (2018). effsize: Efficient Effect Size Computation. R package version 0.7.4, doi: 10.5281/zenodo.1480624

Tóth, A. (2011). Speech disfluencies in simultaneous interpreting: A mirror on cognitive processes. *SKASE Journal of Translation and Interpretation*, *5*(2), 23–31.

Ustaszewski, M., & Stauder, A. (2017). TransBank: Metadata as the Missing Link between NLP and Traditional Translation Studies. In *Proceedings of the Workshop Human-Informed Translation and Interpreting Technology*. Varna, Bulgaria: Association for Computational Linguistics, 29-35.

Van Besien, F., & Meuleman, C. (2014). Style Differences among Simultaneous Interpreters. *The Translator*, *14*(1), 135–155.