SKASE Journal of Translation and Interpretation, 2025; 18(2): 55–78
doi: 10.33542/JTI2025-S-4

Surprisal explains the occurrence of filler particles in simultaneous interpreting

Christina Pollkläsener, University of Hildesheim
Maria Kunilovskaya, Saarland University
Elke Teich, Saarland University

Abstract
We present a set of studies on filler particles in interpreting in the language pair English-German, asking whether, and if so to what extent, surprisal can explain their occurrence. It is widely acknowledged that filler particles are associated with planning effort in monolingual production, but their occurrence in interpreting has not been fully investigated. Surprisal is a measure from information theory that estimates the information content (in bits) of a unit (e.g. a word) in the context of other linguistic units (e.g. n preceding words). Importantly, there is abundant evidence that surprisal is proportional to cognitive effort, measured independently in behavioural experiments, which provides a useful tool for explaining the link between linguistic choice and language cognition. Looking at the words preceding and following a filler particle in the target language output, we investigate surprisal as a predictor of filler particle occurrence, placing a special focus on function words vs. content words that follow filler particles. Overall, our analysis shows that surprisal of following words is a good predictor of filler particles in interpreting.

Keywords: *Filler particles, interpreting, surprisal, prediction, cognitive effort*

1. Introduction

Filler particles (e.g. *euh* or *hum,* also called filled pauses or hesitation markers) are a frequent phenomenon in spoken language. These particles are syntactically nonessential, void of propositional meaning and often linked to difficulties in linguistic planning and production (e.g. Clark & Fox Tree 2002). These properties render them particularly interesting in the context of simultaneous interpreting – a complex task requiring concurrent comprehension of source language input and production of target language output. Previous research has shown that filler particles are overall more frequent in interpreting than in non-mediated (original) speech (e.g. Plevoets & Defrancq 2016, 2018; Chmiel et al. 2022), highlighting their potential role in managing the demands of this dual process. This finding is often linked to variables that are thought to induce an increase in cognitive load such as the presence of numbers, high delivery rate of source speech, interpreting direction (into native or non-native language) and high lexical density (e.g. Plevoets & Defrancq 2016, 2018; Chmiel et al. 2022, 2023, 2024; Bartłomiejczyk & Gumul 2024; Kajzer-Wietrzny et al. 2024). However, to our knowledge, no existing study has examined the immediate linguistic context surrounding filler particles in interpreting. This study addresses this gap by specifically investigating lexico-grammatical properties of the words directly preceding and following filler particles, focussing specifically on their surprisal.

The concept of surprisal, a measure grounded in information theory (Shannon 1948), refers to the (un)expectedness of a given word based on its probability in context. Importantly, surprisal has a direct link to cognition as it is proportional to cognitive effort, measured independently by reading times, eye fixation or event-related brain potentials in comprehension (e.g. Demberg & Keller 2008; Kutas et al. 2011; Smith & Levy 2013). Highly predictable (low surprisal) words are associated with lower cognitive effort and unexpected (high surprisal) words with higher cognitive effort. On the production side, less predictable words are longer in duration and more likely to be articulated with fuller vowels, and vice versa, more predictable units are shortened and phonetically reduced (e.g. Bell et al. 2003; Malisz et al. 2018). Furthermore, some theories of language production suggest that the predictability of words and structures, as reflected in their surprisal, influences their level of pre-activation in the mental lexicon and thereby their ease of lexical access (Kuperberg & Jaeger 2016; Huettig et al. 2022).

Research on monolingual English speech has demonstrated that filler particles tend to appear after low-surprisal words and are followed by high-surprisal ones, suggesting that they may function as cognitive buffers in anticipation of more challenging upcoming material (Dammalapati et al. 2019, 2021; Zámečník 2019). However, the extent to which this relationship holds in the complex, bilingual environment of simultaneous interpreting has not yet been examined.

Focusing on the language pair German-English, the primary goal of our paper is to ask whether surprisal may also explain the occurrence of filler particles in interpreting and what the differences are, if any, comparing free original speech and interpreting. Since interpreters are not fully in control of their own speech and the source language input affects the retrieval processes for the target language output, we assume that surprisal will affect interpreters differently compared to original speakers. As a secondary goal, we pursue a methodological refinement of existing approaches by considering the surprisal of the next content word rather than of the next word. Consider example (1).

(1) speed up euh the spending procedure

Here, the interpreter probably did not struggle to retrieve the function word *the* but rather the following lexical compound *spending procedure*. To test this assumption, we investigate whether the occurrence of filler particles correlates better with the surprisal of the directly following function word or the nearest content word. We think that focusing on the next content word could yield more meaningful results as previous research indicates that there are processing and retrieval differences between content and function words (Bell et al. 2009; Boye & Harder 2012; Lange et al. 2017; Seifart et al. 2018). Filler particles are known to appear at phrase boundaries (e.g. Goldman-Eisler 1968; Schneider 2014; Zámečník 2019) which are often followed by function words such as articles or prepositions. Assuming that high surprisal is also an index of word-planning effort and that filler particles indicate planning difficulty, the surprisal of an upcoming lexical item should be more meaningful compared to the surprisal of a directly following function word, like a preposition or article.

2. State-of-the-Art: Filler particles in interpreting

Filler particles are a specific kind of disfluency in speech production that signals planning effort on the part of the speaker (Corley & Stewart 2008, Kosmala & Crible 2022). In interpreting, filler particles have been shown to be a typical feature along with other markers of online oral production. In fact, features of orality seem to be overemphasised in interpreting when compared to original productions in the same language, as originally shown by Shlesinger & Ordan (2012) for the language pair Hebrew-English and more recently by Przybyl et al. (2022b) using European Parliament data in the language pair English-German, the same data set we use in the present study (see details in Section 3). In this data set, filler particles are more frequent in interpreting compared to original productions in the source language, as is the case in the French-Dutch corpus of Plevoets & Defrancq (2016, 2018) and the Polish-English corpus of Chmiel et al. (2022). In contrast, Dayter (2021) found fewer filler particles in Russian-English interpreting compared to original English speeches. Tissi (2000), Wang & Li (2014) and Bartłomiejczyk & Gumul (2024) found considerable individual variation in filler particle production across interpreters.

For possible explanations of filler particle occurrence in interpreting, researchers have investigated variables that are thought to be linked to increase or decrease in cognitive load. Plevoets & Defrancq (2016, 2018) investigated delivery rate, lexical density, percentage of numerals, average sentence length and formulaicity in the language pair French-Dutch. In their first study (Plevoets & Defrancq 2016), they found a connection between high source text delivery rate, high lexical density in the target text and the frequency of filler particles in interpreted speeches. Source text delivery rate stopped being a significant predictor in their second study (Plevoets & Defrancq 2018), where they used a different statistical method modelling sentence level instead of text level, and introduced formulaicity as a new predictor. In this study, they found a significant effect of source text lexical density and formulaicity of the source and target text. Even though Plevoets & Defrancq (2016, 2018) failed to find a significant effect for percentage of numerals, Kajzer-Wietrzny et al. (2024) took a closer look at numbers as challenging items in interpreting. They distinguished between different types of numbers and included omissions in their model, assuming that interpreters often omit challenging items when cognitive load is too high. In their study, frequency of numbers in the source, type of numbers in the source and omissions in the target were significant predictors of the presence of a disfluency in a sentence. Chmiel et al. (2023) examined interpreters' current load, i.e. the duration of filler particles and silent pauses in the target speech within 500 ms of hearing nouns in the source. They discovered that interpreters paused longer after encountering low-frequency nouns compared to high-frequency nouns. In a later study, Chmiel et al. (2024) found that the mean dependency length of source speeches significantly affected the average pause duration in interpreting but had no impact on the frequency of pauses. In another study on the language pair English-Polish, Bartłomiejczyk & Gumul (2024) started from the assumption that cognitive load is lower for into-A interpreting and compare A-B and B/C-A interpretations from plenary debates of the European Parliament in the language pair English-Polish. They found that interpreters who interpreted into their B-language did not produce a higher rate of filler particles than interpreters who interpreted into their A-language.

These existing works on possible causes for disfluencies in interpreting have in common that they look at quantitative figures over whole texts, (sub)corpora, segments or sentences (e.g. lexical density, delivery rate, interpreting direction), trying to connect aggregated features (e.g. duration of pauses and filler particles, presence of any kind of disfluency) to meta-data that may be indicative of cognitive load. While methodologically perfectly sound, what is hardly considered are the local linguistic conditions of filler particle occurrence, i.e. their position in the clause or sentence or their lexico-grammatical surroundings. Wang & Li (2014) investigated the syntactic distribution of silent pauses and filler particles in Chinese-English interpreting. They found a high proportion of pauses and filler particles inside phrases in simultaneous interpreting, which is marked because they usually appear at phrase boundaries (e.g. Goldman-Eisler 1968; Schneider 2014). However, they did not look further into causes or predictors for this unusual pause and filler particle placement.

In our research, we address these open questions. First, we consider the *local linguistic context* of filler particles; second, we operationalise retrieval difficulty using the concept of *information* as originally defined by Shannon (1948), commonly referred to as *surprisal* (Crocker et al. 2015), i.e. the (un)expectedness of a linguistic unit in context. If a word is highly expected, i.e. very likely to appear in a given context, it provides little new information (low surprisal) and vice versa, if a word is very unlikely in a given context, it provides much information (high surprisal). Surprisal is measured as the negative logarithm of a word's probability given its preceding context (Hale 2001; Levy 2008). This probability is typically derived from a large corpus or language model, though the definition of context may vary from n-gram models that only take into account the directly preceding word to transformer models with context windows of, e.g., 1024 words. Importantly, surprisal is proportional to cognitive effort (Demberg & Keller 2008) and provides a link between linguistic choice and cognitive behaviour. In the psycholinguistics literature, the term *transitional probability* is usually synonymous to surprisal. Recently, surprisal and related measures have also been used to analyse translation (Schaeffer et al. 2015; Teich et al. 2020; Lapshinova-Koltunski et al. 2022).

Surprisal is often used as an operationalisation for prediction or pre-activation of words and structures (for an overview and discussion, see Kuperberg & Jaeger 2016). Studies on prediction in simultaneous interpreting typically approach the topic from the perspective of prediction during comprehension, assuming that high predictability facilitates processing, enabling interpreters to comprehend more efficiently and use predictions to plan their utterances faster (Amos & Pickering 2020). In a study on predictive processes during simultaneous interpreting from German into English, Hodzik & Williams (2017) examined interpreters' latency times for interpreting sentence-final verbs in the source language input. They found an impact of contextual constraints on prediction but not of transitional probability (surprisal). However, their calculation of transitional probability was based on an n-gram model only considering the directly preceding word. Given its broad context window, transformer surprisal may have captured both the probability of the verb based on the preceding word and the influence of contextual constraints earlier in the sentence. Amos et al. (2022) found evidence for the prediction of upcoming nouns in simultaneous interpreting in a visual world paradigm study. While interpreting sentences with a highly predictable word from English into French, participants made predictive eye movements towards a picture of that word before it was spoken in the English sentence. Liu et al. (2022) also found that 75% of their participants made predictive eye movements in a visual world paradigm while engaging in simultaneous interpreting. However, they found that 25% of their participants did not show predictive eye movements, which they speculated to be the result of the extreme cognitive burden during simultaneous interpreting, which could be a limit on prediction (see also Huettig & Guerra 2019; Amos & Pickering 2020).

In the study at hand, we look at a different angle of prediction in simultaneous interpreting. We focus on prediction and its role in lexical access in language production and planning, that is, interpreters' production and planning of the target language directly. In this sense, low surprisal is tantamount to high pre-activation of the word and therefore ease in lexical access, whereas high surprisal means low pre-activation, leading to more difficulty in lexical access (for a detailed explanation of this process, see Huettig et al. 2022). Greater difficulty in lexical access should then be associated with higher filler particle occurrence, as filler particles signal difficulties in speech planning (Clark & Fox Tree 2002; Corley & Stewart 2008; Kosmala & Crible 2022).

Early works on filler particle placement in monolingual settings (Goldman-Eisler 1968; Beattie & Butterworth 1979) already hypothesised that there is a connection between predictability and filler particles, but these early studies were lacking in methodological sophistication. Some more recent studies have shown that disfluencies occur in the presence of production difficulties due to new information, where "new" could well be modelled with surprisal (e.g. Barr 2001; Arnold et al. 2003; Heller et al. 2015). Investigating monolingual English, Dammalapati et al. (2019, 2021) found that disfluencies (filler particles and repairs) tend to appear after low surprisal words and before high surprisal words. They interpreted this as speakers selecting easily accessible words (low surprisal) to better plan for upcoming production difficulties, as reflected in high surprisal words following disfluencies. Zámečník (2019) also identified surprisal as a strong predictor for filler particle occurrence in monolingual English speech.

Here, we build on these insights from monolingual language production and apply it to simultaneous interpreting. We assume that at least part of lexical access in interpreting is similar to that of monolingual speech production. However, we know that cross-lingual lexical activation happens (Amos & Pickering 2020). Furthermore, Amos & Pickering (2020: 710) note that cognitive load could hinder prediction. More specifically, Amos et al. (2022: 3) point out that "concurrent activation of both languages, and the regular switching of focus from comprehension to production and back between the two languages, may lead to a weaker activation of each language". Therefore, we expect differences in the effect of surprisal on filler particle occurrence between original and interpreted speeches.

3. Methodology

3.1. Data

The data used in this study are drawn from the German and English subcorpora of EPIC-UdS (Przybyl et al. 2022a). It contains transcriptions of speeches held at the European Parliament between 2008 and 2013 by English and German native-speaking MEPs and their interpretations into German and English, respectively. The transcriptions include disfluencies such as filler particles, truncations and repetitions. The corpora include rich annotations such as part of speech, dependency relations and lemma. Speaker and interpreter identity are included in the metadata. A student assistant was asked to identify individual interpreters by listening to interpreters' voices. Table 1 shows the size of the individual subcorpora. The total corpus comprises approximately 250,000 tokens.

Table 1: EPIC-UdS Corpus (English and German subcorpora)

Subcorpus	Tokens
English Interpreting (SI)	58,503
English Originals (ORG)	71,146
German Interpreting (SI)	62,120
German Originals (ORG)	56,488

3.2. Processing and annotation

Surprisal is typically estimated with computational language models based on large corpora (e.g. n-gram models, LSTMs, Transformers such as Llama or GPT). In our data, surprisal for all tokens in the English corpora was annotated using the default, monolingual GPT-2(-small) model (Radford et al. 2019). There is evidence that GPT-2-small achieves a substantially better fit to reading times than larger GPT-2 models (e.g. GPT-2-XL) or more recent GPT models (e.g. GPT-3), leading to its selection over these alternatives (Oh & Schuler 2021; Shain 2024). For the German data, the corresponding German GPT-2 model (Schweter 2020) was used. Prior to surprisal indexing and analysis, the data were parsed with Stanza (Qi et al. 2020), and surprisal values were calculated for tokens as defined by the parser. Hyphenated words were treated as one token.

Filler particles were removed before parsing and surprisal annotation and reinserted into the corpus afterwards. The rationale for this was that otherwise they would have influenced the surprisal values for tokens following the filler particle as they would have been considered part of the prior context of these words. We did not distinguish between filler particle realisations with and without nasal consonants (*euh hum hm*). Clusters (*euh euh*) were counted as a single instance of a filler particle occurrence.

In addition to surprisal, word frequency and word length were annotated for each word in the corpus. In psycholinguistic literature, both variables are empirically shown to influence language processing, impacting measures such as reading times (e.g. Brysbaert et al. 2017; Kuperman et al. 2024) and event-related potentials (Hauk & Pulvermüller 2004). To disentangle the effect of frequency and length from any possible effect of surprisal, these variables were included in the statistical analysis. Word length was operationalised as number of characters per word and frequency was annotated using frequency measures from SUBTLEX-UK (Van Heuven et al. 2014) and SUBTLEX-DE (Brysbaert et al. 2011).

Since we were interested in investigating differences in lexical contexts where filler particles are more or less likely to occur, for every filler particle, the surprisal, length and frequency of the directly preceding and following word was annotated. Since filler particle can potentially be freely inserted between any two words, any place between two words in the corpus where a filler particle did not appear in was considered a potential filler particle placement and annotated with the surprisal, frequency and length values of the preceding and following words. We controlled for sentences. This means that filler particles and potential filler particle placements at the very beginning or at the very end of sentences were excluded from the study. Example 2 shows our annotation in practice: Example (2b) is an instance of a filler particle occurrence, and examples (2a), (2c) and (2d) are instances of slots where a filler particle could have potentially occurred but did not. Filler particles can also appear within words. However, those cases are rather infrequent (e.g. Defrancq & Plevoets 2018), which is why we decided to focus on the potential placement between words.

(2) a. under (8.41 bits) **Ø [0]** the (3.04 bits) EUH Posted Workers Directive

b. under the (3.04 bits) **EUH [1]** Posted (25.76 bits) Workers Directive

c. under the EUH Posted (25.76 bits) **Ø [0]** Workers (11.06 bits) Directive

d. under the EUH Posted Workers (11.06 bits) **Ø [0]** Directive (2.68 bits)

For Study 2, we were interested in the surprisal value of the next following content word. We considered adjectives, adverbs, nouns, proper nouns, verbs and numerals as content words and conjunctions, particles, determiners, prepositions, auxiliaries and pronouns as function words. For both actual and potential filler particle placements, we annotated the surprisal of the next following content word, which is seen in examples (3a–c). Cases where the directly following word was a content word, such as (3a) and (3c), were deleted for the analysis in Study 2.

(3) a. **Ø [0]** statement ((4.71)) EUH in Vietnam

b. statement **EUH [1]** in Vietnam ((14.90))

c. statement EUH in **Ø [0]** Vietnam ((14.90))

3.3. Statistical analysis

To examine the relationship between lexical surprisal and the occurrence of filler particles, a mixed-effects logistic regression model was employed. Table 2 shows an overview of all variables in the models. The dependent variable was binary, indicating the presence or absence of a filler particle (1-present filler particle; 0-absent filler particle) at each annotated position. As described in Section 3.2., primary predictors were the surprisal values of the preceding and following words, with both variables included as fixed effects. Additionally, word frequency and word length were included as covariates to adjust for their possibly confounding influence on language processing. Variance Inflation Factor (VIF) values were calculated to assess multicollinearity among continuous predictors (cf. Table 2). All VIF values were below the commonly accepted threshold of 10 (Montgomery & Peck 1992), indicating that multicollinearity was not a concern and that the predictors could be reliably interpreted.

Since both original and interpreted speeches are included in the data and we wanted to investigate whether surprisal affects filler particle occurrence in interpreted language differently to original, monolingual speeches, we included the variable of mode. Mode indicates if the present or absent filler particle stems from the interpreted corpus or the original one. Mode is a categorical variable that has been sum-coded (1-interpreted speeches; -1-original speeches). Under this coding scheme, the main effects represent the average effect across both levels of the mode variable, effectively reflecting a midpoint between original and interpreted speeches.

Furthermore, we fitted interactions between mode and surprisal, frequency and length. The interactions indicate if there is a difference in how surprisal, frequency and length affect filler particle occurrence in interpreting and original speeches. Random intercepts were incorporated for speakers and interpreters to account for variability attributable to individual differences.

We fitted separate statistical models for the German and the English data. This decision was motivated by the fact that while surprisal and frequency values for both languages were derived from comparable corpora or language models, they were not identical. Corpus size and morphological differences between the languages affect the calculation of both surprisal and frequency. A single model for both languages would have treated these values as identical and calculated a single estimate. Fitting separate models prevents a direct comparison of estimates but avoids potential biases.

All continuous variables were z-scored and centered to make them comparable amongst each other (see Winter 2019: 86–89). To account for the skewed distribution typically associated with frequency data, all frequency values were log-transformed prior to analysis. As the regession model is a logistic one, all coefficients are expressed in terms of log-odds. All statistical analyses were conducted in R (R Core Team 2024) using the car (Fox & Weisberg 2019), ggplot2 (Wickham 2016) and lme4 packages (Bates et al. 2015).

Table 2: Summary of variables included in the regression models for English and German data, including their type, description, and Variance Inflation Factor (VIF) values

Variable	*Variable type, description*	*VIF English model*	*VIF German model*
Filler particle	Response variable, binary (1-filler particle, 0-no filler particle)	N/A	N/A
Preceding surprisal	predictor, continuous gpt2 estimation	1.503851	1.241809
Following surprisal	predictor, continuous gpt2 estimation	3.342427	2.244370
Next content word surprisal	predictor, continuous gpt2 estimation	3.262839	2.208978
Preceding length	predictor, continuous character length	2.530324	3.154531
Following length	predictor, continuous character length	5.186765	5.866501
Next content word length	predictor, continuous character length	3.249556	3.963820
Preceding frequency	predictor, continuous SUBTLEX	3.089237	3.245423
Following frequency	predictor, continuous SUBTLEX	5.930943	5.488847
Next content word frequency	predictor, continuous SUBTLEX	3.887907	3.489127
Mode (ORG vs. SI)	predictor, categorical metadata in EPIC-UdS (1 – SI, -1 – ORG)	N/A	N/A
Speaker or interpreter ID	random effect, categorical metadata in EPIC-UdS	N/A	N/A

4. Study 1 – following and preceding surprisal

In this section, we focus on the key predictors of preceding and following surprisal as well as their interactions with mode (original vs. interpreted speech). Additional predictors were included in the models solely for adjustment purposes, to account for potential confounding effects.

4.1. Results

4.1.1. English model

Table 3 shows the results for the fixed effects of the logistic regression model for the English data and Figure 1 depicts the main effects of following and preceding surprisal on filler particle occurrence. Focussing first on the main effects in Table 3, we see that both preceding and following surprisal are significant. By adjusting for length and frequency in the model, the observed effect appears to originate genuinely from surprisal, free from the influence of potential confounding variables. The negative estimate for preceding surprisal suggests that as surprisal values increase, the likelihood of the next word being a filler particle decreases, which can be seen in Figure 1. This implies that words preceding filler particles tend to have lower surprisal values.

Focussing on following surprisal, its positive estimate indicates that the higher the surprisal of the following word, the more likely it is preceded by a filler particle, as shown by the rising slope in Figure 1. This implies that filler particles often occur before less predictable lexical items. Furthermore, following word surprisal has the highest estimate among the continuous predictors, highlighting its important role in the occurrence of filler particles.

Examining the interactions, the data reveals that there is no statistically significant difference between interpreting and originals for preceding surprisal, nor for following word surprisal. Both effects on filler particle usage remain consistent regardless of whether the speech is interpreted or original. The estimate for mode reveals that filler particles are more prevalent in interpreting compared to original speeches.

Table 3: Parameter estimates for the English data for the fixed effects. Significant effects are marked with an asterisk (*)

Predictor	*Estimate*	*Standard error*	*Z value*	*p-value*
Intercept	-4.09711	0.12410	-33.016	<0.0001*
Preceding surprisal	**-0.14375**	**0.02416**	**-5.950**	**<0.0001***
Following surprisal	**0.50922**	**0.02247**	**22.658**	**<0.0001***
Preceding length	0.21973	0.02810	7.820	<0.0001*
Following length	0.12503	0.02826	4.424	<0.0001*
Preceding frequency	-0.10932	0.03224	-3.391	<0.0001*
Following frequency	0.21277	0.03447	6.173	<0.0001*
Mode (ORG vs. SI)	0.56474	0.12343	4.575	<0.0001*
Preceding surprisal x mode	**0.01342**	**0.02416**	**0.555**	**0.57855**
Following surprisal x mode	**0.01653**	**0.02247**	**0.736**	**0.46190**
Preceding length x mode	-0.04755	0.02810	-1.692	0.09062
Following length x mode	-0.0186	0.02826	-0.658	0.51037
Preceding frequency x mode	0.06542	0.03224	2.029	0.04245*
Following frequency x mode	-0.09876	0.03447	-2.865	0.00416*

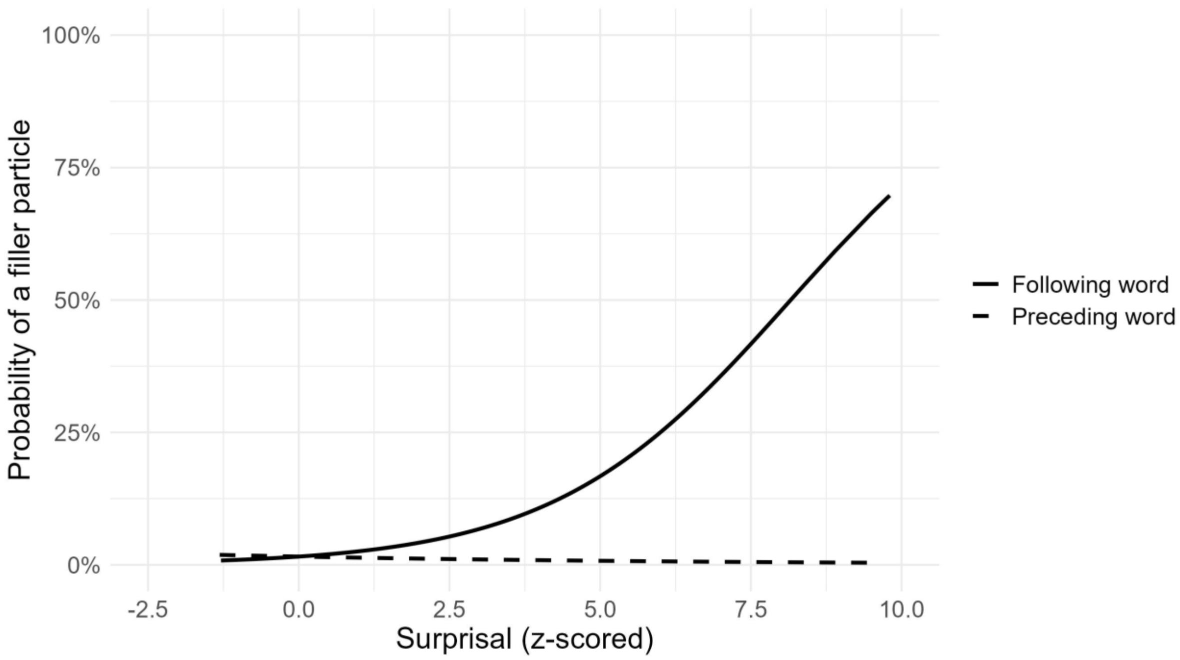

Figure 1: Effect of preceding and following word surprisal on filler particle occurrence in English, surprisal values are z-scored

4.1.2. German model

Table 4 shows the results for the fixed effects for the logistic regression model for the German data, and Figure 2 depicts the effects of following and preceding surprisal and their interactions with mode. Focussing first on the main effects, we see that, in contrast to the results for the English data, preceding surprisal is not significant and has a positive estimate, which is not the direction we expected. The results for following surprisal are more similar to the English results in Section 4.1.1. It is significant and has the highest, positive estimate among the continuous predictors. This result indicates that higher surprisal values for following words strongly increase the probability of a filler particle occurring. The estimate for mode reveals that filler particles are substantially more prevalent in German interpreting compared to German original speeches. This big difference can be observed in the gap between the two intercepts at 0 in Figure 2.

Examining the interactions, the German data differs from the English results. The interaction between preceding surprisal and mode is significant, indicating that the relationship between preceding surprisal and filler particle occurrence varies slightly between interpreted and original speeches. However, the main effect of preceding surprisal is not significant in the first place. Furthermore, since the main effect of preceding surprisal is very low, taking the interaction into account, the interaction term, preceding surprisal has a positive estimate for interpreting (0.00011 + 0.05362 = 0.05373) and a negative estimate for originals (0.00011 − 0.05362 = −0.05351). Figure 2 shows that the lines for preceding word surprisal are nearly straight, indicating that any effect, regardless of its direction, is minimal. The interaction between following surprisal and mode in the German data is also significant. It has a negative estimate, indicating that the effect of following surprisal is more pronounced in German original speeches compared to interpreted ones. This can be observed in the steeper slope for originals (red line) compared to interpreting (blue line) in Figure 2, even though the overall propability of filler particle occurrence is still higher in interpreting.

Table 4: Parameter estimates for the German data for the fixed effects. Significant effects are marked with an asterisk (*)

Predictor	*Estimate*	*Standard error*	*Z value*	*p-value*
Intercept	-4.17154	0.12543	-33.257	<0.0001*
Preceding surprisal	**0.00011**	**0.02585**	**0.004**	**0.99648**
Following surprisal	**0.41818**	**0.02183**	**19.154**	**<0.0001***
Preceding length	0.09575	0.03714	2.578	0.00993*
Following length	-0.02101	0.03945	-0.533	0.59431
Preceding frequency	-0.09245	0.04003	-2.309	0.02092*
Following frequency	0.18933	0.04277	4.427	<0.0001*
Mode (ORG vs. SI)	1.10942	0.12488	8.884	<0.0001*
Preceding surprisal x mode	**0.05362**	**0.02585**	**2.074**	**0.03808***
Following surprisal x mode	**-0.08705**	**0.02183**	**-3.987**	**<0.0001***
Preceding length x mode	-0.02183	0.03714	-0.588	0.55662
Following length x mode	-0.04493	0.03945	-1.139	0.25472
Preceding frequency x mode	0.07754	0.04003	1.937	0.05273
Following frequency x mode	-0.14936	0.04277	-3.492	0.00048*

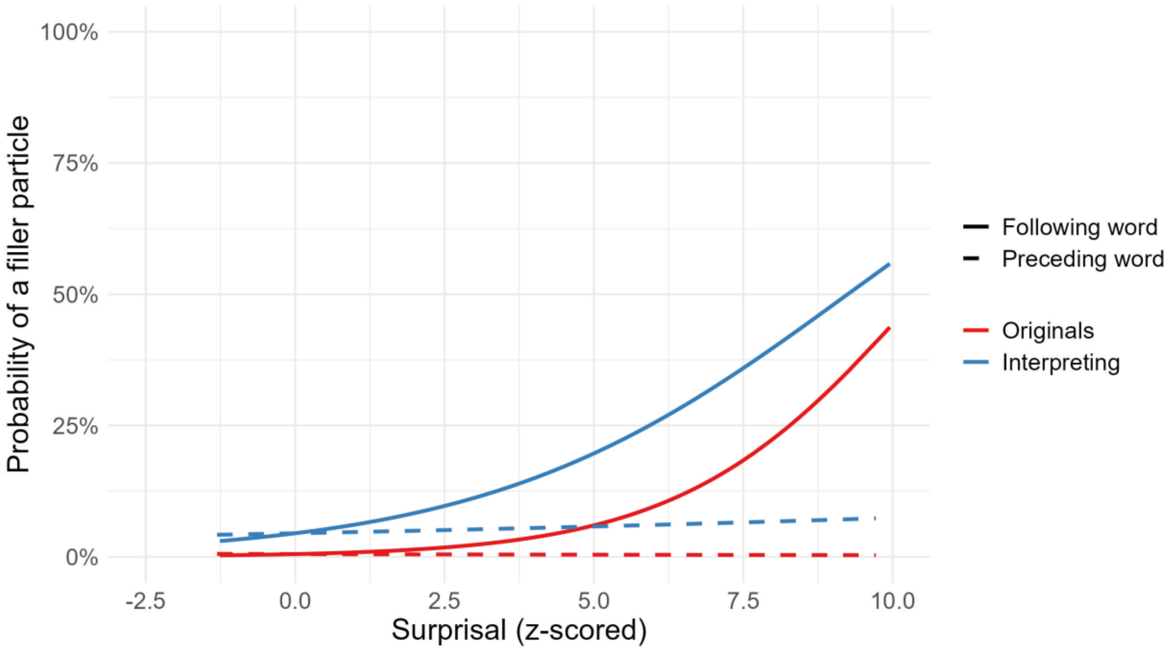

Figure 2: Effect of preceding and following word surprisal on filler particle occurrence in German, grouped by mode (original vs. interpreting), surprisal values are z-scored

4.2. Discussion

Overall, the findings confirm that the (un)predictability of the surrounding words (as captured with surprisal) influences filler particle production in both German and English interpreting, extending findings from previous research on monolingual English to bilingual speech production in interpreting. More specifically, this result is more pronounced for following surprisal than for preceding surprisal. Preceding surprisal is significant for the English data, replicating Dammalapati et al.'s (2019, 2021) findings. They interpreted the lowering in surprisal before disfluencies as a strategy of the speaker to free up resources for upcoming speech planning difficulties. However, we do not see this effect in the German data. This could hint at cross-linguistic differences and should be investigated in more languages.

Following word surprisal consistently demonstrates the strongest effect across both languages, with the highest estimates among all predictors, underscoring its significant role in filler particle occurrence. This suggests that high surprisal words are at a planning disadvantage compared to low surprisal words, as their low predictability from the preceding context results in less pre-activation and more planning difficulty, as indicated by filler particle occurrence. Huettig et al. (2022) describe pre-activation through predictability in speech planning as a form of pattern completion: Uttering the beginning of a sentence or phrase activates possible continuations. The more constraining the context (lower surprisal), the greater the activation of lexical items and the easier lexical access. Conversely, the less constraining the context (higher surprisal), the lesser the activation of lexical items and the more taxing lexical access, as reflected in the higher probability of filler particle occurrence.

Taken together with findings from Hodzik & Williams (2017), Amos et al. (2022) and Liu et al. (2022), our findings suggest that prediction in simultaneous interpreting takes place in both comprehension and speech planning, leading to activations in both source and target language. This could mean activation of more entries in both languages and therefore weaker activation of these entries (Amos & Pickering 2020), which could explain why surprisal has a weaker effect on filler particle occurrence for interpreters compared to original speakers in our German results. However, since this difference was not found for English, the effect of surprisal and prediction in simultaneous interpreting may be language or language pair-specific and therefore should be investigated in more languages.

Table 5: List of most frequent words following filler particles, grouped by mode and language

Original English	*Frequency per 1,000 tokens*	*English Interpret ing*	*Frequency per 1,000 tokens*	*Original German*	*Frequency per 1,000 tokens*	*German Interpret ing*	*Frequency per 1,000 tokens*
the	46.39	the	35.04	die	22.74	die	24.50
and	25.73	to	21.01	der	19.75	und	20.94
to	25.68	and	20.05	und	19.23	der	15.34
of	23.62	of	16.92	in	14.99	in	13.01
that	18.05	that	13.61	dem	9.55	das	12.49

Since the surprisal of the following word seemed to have a considerable effect on filler particle production in both German and English, we decided to take a closer look at the following context of filler particles. Table 5 shows that the most frequent words following filler particles are articles, prepositions and conjunctions. This is in itself not surprising since filler particles are known to appear at phrase boundaries (e.g. Goldman-Eisler 1968; Schneider 2014). However, literature suggests differences in retrieval of function and content words (Bell et al. 2009; Boye & Harder 2012; Lange et al. 2017; Seifart et al. 2018) and surprisal of function words tends to be lower than surprisal of content words. Since the retrieval of function words is probably not the problem but rather the retrieval of the next upcoming content word, (see

Example 3), we conducted a second study where we compared the surprisal of the next upcoming content word as a predictor of filler particles to the surprisal of the directly following function word.

5. Study 2 – following and preceding surprisal

In Section 5.1., the analysis focuses on the comparison between surprisal of the next content word and surprisal of the directly following function word as predictors for filler particle occurrence. To do this, we create a model that only contains words and filler particles that are followed by function words (conjunctions, particles, determiners, prepositions, auxiliaries and pronouns) and compare the coefficients of the directly following function word and the next content word. As explained in Section 3.2., occurrences of filler particles and potential placements directly followed by a content word were excluded. Furthermore, we conduct model comparisons, comparing the full model with both surprisal measures, once to a model without directly following function word surprisal and once to a model without next content word surprisal. Additional control variables (preceding surprisal, length, frequency) are shown in the results table for the sake of completeness but will not be addressed specifically.

5.1. Results

5.1.1. English model

Table 6 shows the results for the fixed effects of the full model with both following word surprisal predictors along with control variables for the English data, and Figure 3 depicts the effects of both following word surprisal predictors and their interactions with mode. Different to what we expected, both directly following word surprisal and next content word surprisal significantly influence the likelihood of filler particle occurrence. Focussing on the estimates, we see that the surprisal of the directly following word has a stronger positive effect on filler particle occurrence than the surprisal of the next following content word. This can be seen in the steeper slope for directly following word surprisal in Figure 3. Shrout & Yip-Bannicq (2017) highlight that traditional regression model outputs do not indicate whether two coefficients differ significantly from one another, underscoring the importance of conducting specific tests (see also Gelman & Stern 2006). We tested the difference between the two coefficients using the linearHypothesis() function of the car package in R (Fox & Weisberg 2019). This revealed a significant difference between the two following surprisal estimates ($p < 0.0001$). The direction of the difference runs counter our expectations since we expected the surprisal of the next content word to be a more meaningful predictor of filler particles than the surprisal of the directly following word, but it is the other way around.

Next, we discuss the model comparisons. As we expected, a likelihood ratio test of the model including both surprisal measures against a model without next content word surprisal reveals a significant difference between the models ($\chi^2(1) = 17.922$, $p < 0.001$), indicating that next content word surprisal adds relevant information to the model. However, against our expectations, a likelihood ratio test of the model including both surprisal measures against the model without directly following word surprisal also reveals a significant difference between the models ($\chi^2(1) = 147.17$, $p < 0.0001$). This indicates that both surprisal measures of the following context of filler particles contribute uniquely to the model and have separate effects. Notably, this outcome contradicts our initial expectation, as we assumed that both measures reflect the same underlying phenomenon but that content word surprisal would emerge as the superior measure.

To explain these results, we take a closer look at the distributions of the surprisal values of the directly following words per word class (cf. Figure 4). Comparing content and function words, it is apparent that content words have higher surprisal values in general. Among content words, those following filler particles display notably higher surprisal values compared to those that do not follow filler particles. This aligns with our expectations. However, a similar trend emerges for function words: both the box and median surprisal values are higher for function words following filler particles than for those not following them. This could explain the regression results as it seems like that, regardless of word class, words following filler particles have higher surprisal.

Table 6 also shows that neither the interaction between mode and directly following function word surprisal nor the interaction between mode and next content word surprisal is significant. This indicates there is no difference in the effect in interpreting compared to original speeches

Table 6: Parameter estimates for the English data for the fixed effects; significant effects are marked with an asterisk (*)

Predictor	*Estimate*	*Standard error*	*Z value*	*p-value*
Intercept	-4.39793	0.13302	-3.795	<0.0001*
Preceding surprisal	-0.16780	0.04422	13.297	0.00015*
Directly following word surprisal	**0.45792**	**0.03444**	**4.349**	**<0.0001***
Next content word surprisal	**0.19421**	**0.04466**	**5.221**	**<0.0001***
Preceding length	0.28867	0.05529	4.266	<0.0001*
Directly following length	0.19944	0.04675	1.174	<0.0001*
Next content word length	0.05178	0.04411	0.231	0.24039
Preceding frequency	0.01217	0.05266	-1.753	0.07961
Directly following frequency	-0.11049	0.06303	5.805	<0.0001*
Next content word frequency	0.31231	0.05380	3.364	0.81726
Mode (ORG vs. SI)	0.43638	0.12973	-1.624	<0.0001*
Preceding surprisal x mode	-0.07182	0.04421	0.603	0.1043
Directly following surprisal x mode	**0.02075**	**0.03444**	**0.478**	**0.5467**
Next content word surprisal x mode	**0.02134**	**0.04466**	**-1.845**	**0.63278**
Preceding length x mode	-0.10201	0.05529	0.605	0.06504
Directly following length x mode	0.02829	0.04675	1.076	0.54507
Next content word length x mode	0.04747	0.04411	-0.767	0.28184
Preceding frequency x mode	-0.04837	0.06303	-0.007	0.44284
Directly following frequency x mode	-0.00038	0.05380	0.552	0.99438
Next content word frequency x mode	0.02907	0.05265	-3.795	0.58093

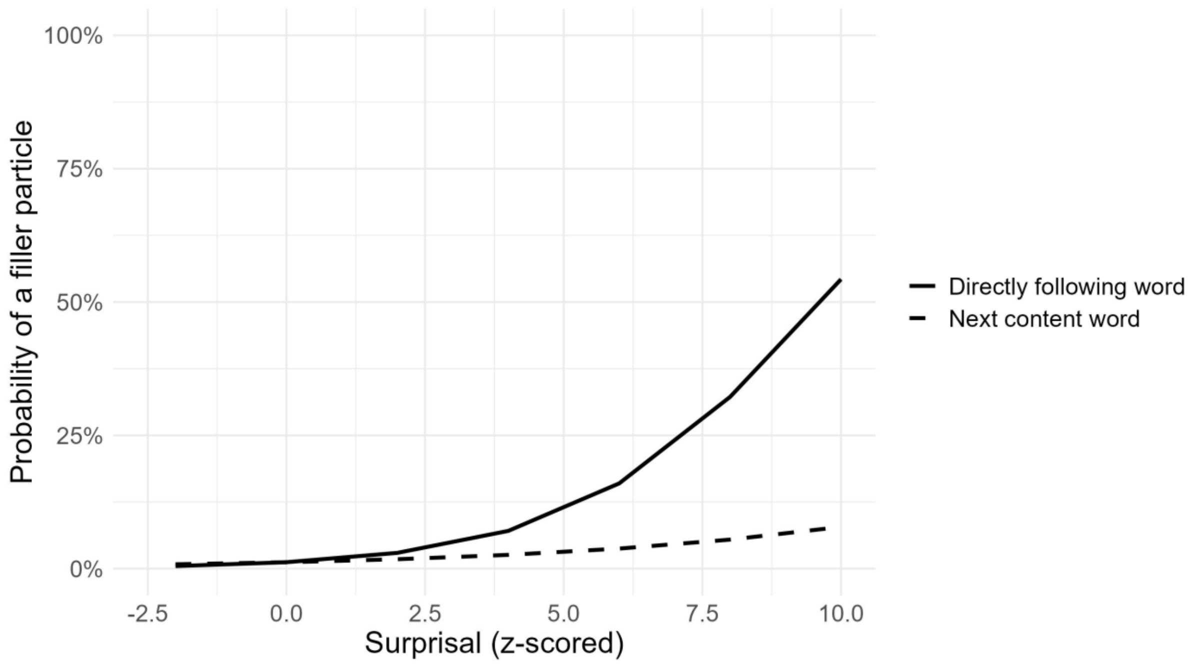

Figure 3: Effect of directly following word and next content word surprisal on filler particle occurrence in English, surprisal values are z-scored

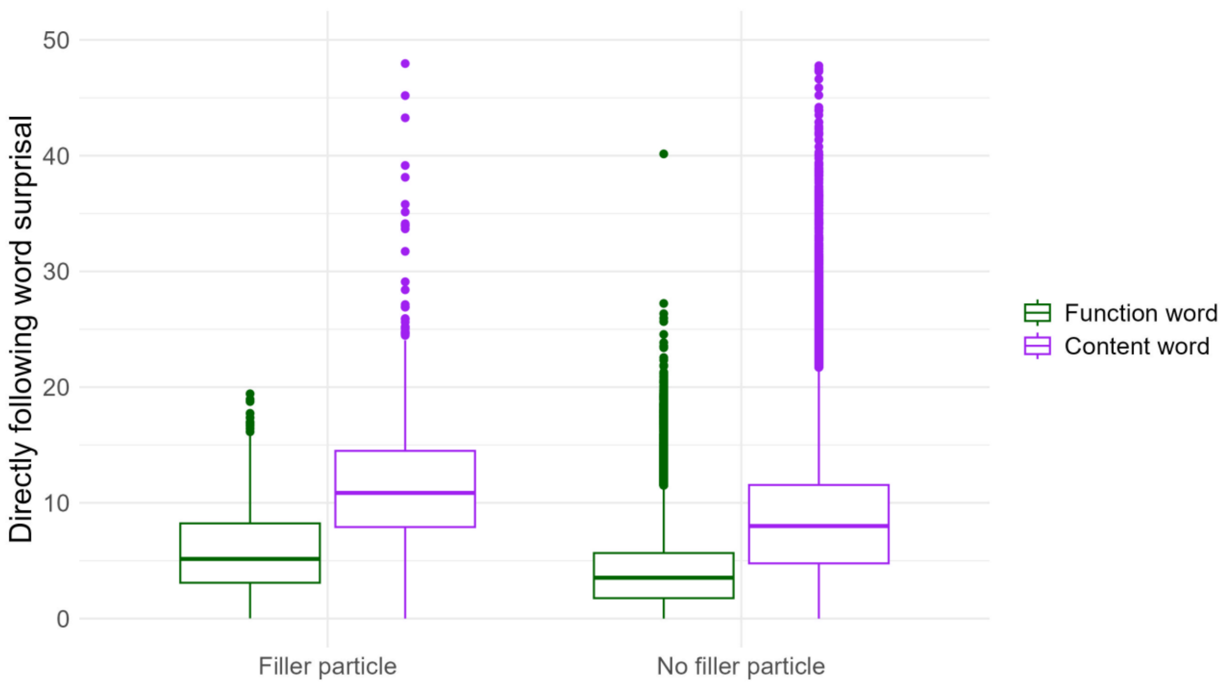

Figure 4: Distribution of surprisal (in bits) of words directly following filler particles or following other words that are not filler particles, grouped by word class (English)

5.1.2. German model

Table 7 and Figure 5 show the results for the main effects for Study 2 on the German data. Again, the model defies our expectations as both surprisal of the directly following word and of the next content word exhibit a significant effect on filler particle occurrence. The effect of directly following word surprisal is again greater compared to next content word surprisal, which can be observed in Figure 5. This time, however, the difference between the two estimates is not significant ($p = 0.2393$), indicating that none of the predictors has a stronger effect on filler particle occurrence compared to the other. There is no significant interaction with mode for neither of the two surprisal measures.

We conducted likelihood ratio tests to compare the model including both surprisal measures with models excluding either next content word surprisal or directly following word surprisal. Consistent with the results for the English data but contrary to our expectations, both comparisons were significant: excluding next content word surprisal ($\chi^2(1) = 22.769$, $p < 0.0001$), and excluding directly following surprisal ($\chi^2(1) = 51.536$, $p < 0.0001$), each revealed significant differences between the models. These findings suggest that each surprisal measure contributes uniquely to the model.

The boxplots in Figure 6 show a similar picture as the boxplots of the English data. Content words have higher surprisal values overall compared to function words. Content words following filler particles have higher surprisal values than content words following other words in the corpus. The same is true for function words, which tend to have higher surprisal values when they appear after filler particles compared to their occurrence after other words. This could explain the significant effect of directly following word surprisal, which only includes surprisal measures of function words.

Table 7: Parameter estimates for the German data for the fixed effects, significant effects are marked with an asterisk (*)

Predictor	*Estimate*	*Standard error*	*Z value*	*p-value*
Intercept	-4.16932	0.14563	-28.630	<0.0001*
Preceding surprisal	-0.07822	0.04856	-1.611	0.10723
Directly following word surprisal	**0.29565**	**0.03723**	**7.942**	**<0.0001***
Next content word surprisal	**0.22495**	**0.04517**	**4.981**	**<0.0001***
Preceding length	0.12163	0.07315	1.663	0.09636
Directly following length	-0.17464	0.06169	-2.831	0.00464
Next content word length	-0.07827	0.05514	-1.420	0.15573
Preceding frequency	-0.01724	0.05925	-0.291	0.80350
Directly following frequency	-0.00095	0.07671	-0.012	0.99013
Next content word frequency	0.13708	0.06295	2.178	0.02943
Mode (ORG vs. SI)	1.1248	0.14258	7.889	<0.0001*
Preceding surprisal x mode	0.08347	0.04856	1.719	0.04187*
Directly following surprisal x mode	**-0.01444**	**0.03723**	**-0.388**	**0.0856**
Next content word surprisal x mode	**-0.02329**	**0.04517**	**-0.516**	**0.69818**
Preceding length x mode	-0.06659	0.07315	-0.910	0.60605
Directly following length x mode	-0.01548	0.06169	-0.251	0.36264
Next content word length x mode	0.09623	0.05514	1.745	0.80185
Preceding frequency x mode	-0.09118	0.07671	-1.189	0.08095
Directly following frequency x mode	-0.05503	0.06295	-0.874	0.23458
Next content word frequency x mode	-0.00982	0.05925	-0.166	0.38199

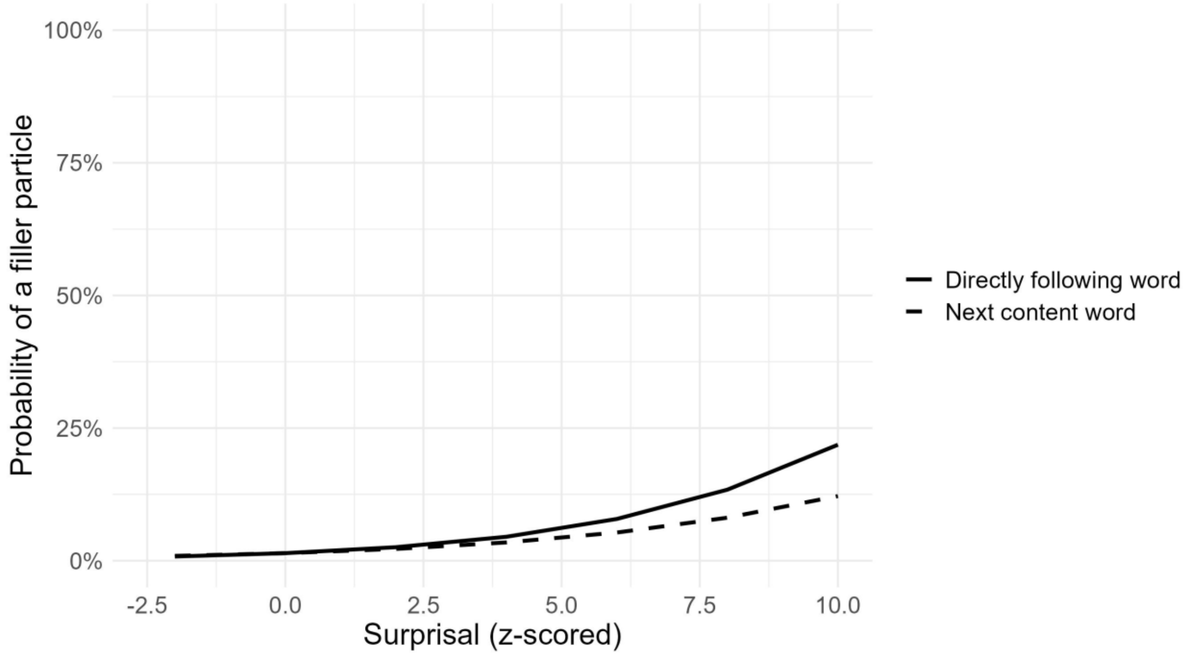

Figure 5: Effect of directly following and next conent word surprisal on filler particle occurrence in German, surprisal values are z-scored

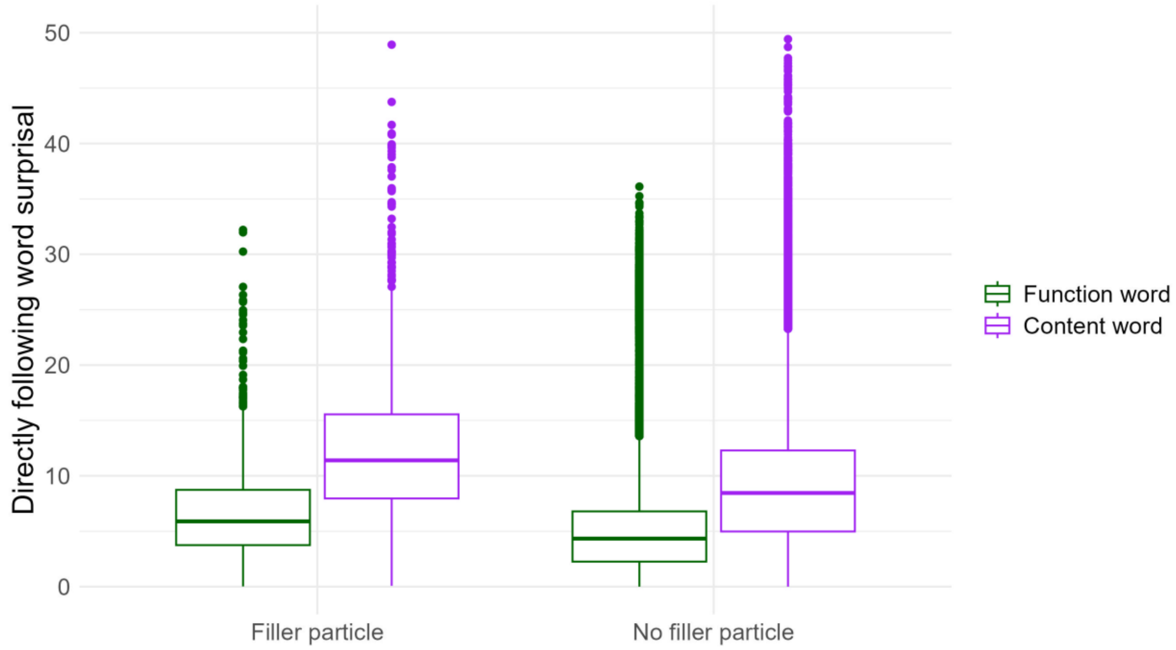

Figure 6: Distribution of surprisal (in bits) of words directly following filler particles or other words that are not filler particles, grouped by word class (German)

5.2. Discussion

Contrary to our initial expectations, next content word surprisal did not arise as the superior measure over directly following function word surprisal. In cases where a filler particle occurs before a funciton word, both the surprisal of the directly following function word and the surprisal of the next content word contribute independently to the occurrence of filler particles. This implies that they might capture distinct aspects of linguistic processing. High surprisal of content words could indicate low lexical activation of the content words themselves in specific contexts, leading to retrieval difficulty and filler particle occurrence. In contrast, high surprisal of function words could reflect speech planning difficulties at a structural level.

Slaats & Martin (2025) argue that while surprisal does not encode syntactic structure, surprisal values are affected by syntactic structure and contain some structural information. For example, a high surprisal preposition may indicate that a prepositional phrase is unexpected in the given context, possibly causing planning difficulty or lesser activation on the structural level. In contexts like example (1), the surprisal of *the* then indicates the unexpectedness of the noun phrase and the surprisal of *spending procedure* indicates the unexpectedness of the lexical item itself. This fits with Huettig et al.'s (2022) parallel architecture pre-activation theory, as they assume an extended lexicon where syntactic structures are stored alongside words and are also pre-activated and accessed in the same way.

Some models of speech production assume that content and function words are accessed at different stages in the production process. According to Boye & Harder (2012), function words are planned later than content words because they typically depend on a lexical head (e.g. a determiner on a noun). The retrieval and encoding of function words rely on the partial processing of their associated lexical head, necessitating a later planning stage for function words compared to content words. This dependency could explain why both the directly following function word surprisal and the next content word suprisal are significant predictors of filler particle occurrence.

Our results suggest that the influence of surprisal in the following context remains consistent between simultaneous interpreting and original speeches, as we find no significant interaction between mode and either of the following surprisal measures. This stands in contrast to the significant interaction between mode and following surprisal in German in Study 1 (see Section 4.1.1.), where following surprisal had a weaker effect in interpreting than in original speeches. Notably, in Study 1, the directly following surprisal measure included both content and function words, whereas in Study 2, it is limited to function words. This suggests that in German, structural access, indicated by surprisal of directly following function words, is less affected by the constraints of interpreting compared to content word access, indicated by surprisal of directly following content words.

6. Summary and conclusion

In this article, we have investigated the linguistic-contextual conditions of filler particle occurrence in simultaneous interpreting in the language pair English-German (both translation directions) using data from European Parliament interactions. Specifically, we posed the question whether, and if so to what extent, filler particle occurrence can be explained with surprisal, i.e. the (un)expectedness of linguistic units in a given context (here: in the target output). Using logistic regression, we conducted two studies that investigated the surprisal values surrounding filler particles in original and interpreted speeches. We were able to show that a higher surprisal of upcoming words increases the probability of a filler particle to occur in interpreting. We saw this as an indication for prediction and pre-activation in the target language and suggested that filler particles indicate retrieval difficulty of upcoming words. Contrary to our initial assumption, we found a surprisal effect both for lexical and function words. This suggests that the occurrence of filler particles may be associated with difficulties in lexical retrieval as well as structural planning of the target language output.

Our future work will therefore involve investigating which structural aspects are likely to play a role in fluent interpreting. For instance, Jiang & Jiang (2020) show that disfluencies may arise due to long dependency distances in the source language input (possibly creating similar processing problems as garden-path sentences, where a unit in processing cannot be concluded because of multiple embedding). Strategies to cope with such situations result in syntactic simplification (see, e.g., Xu & Liu 2023), which, in turn, has been shown to be indexed by surprisal or entropy (Martínez & Teich 2017; Teich et al. 2020; Lapshinova-Koltunski et al. 2022). Here, again focusing on the target language output, we consider taking into account difficulties potentially associated with the surprisal of the syntactic or semantic head of the following construction as well as the complexity of the whole construction.

Finally, we may use surprisal to explain other kinds of phenomena associated with high effort in interpreting. For instance, translation latency, which signals that there is on-going processing of the source language input (see e.g. Chmiel 2021), may well be assumed to be linked to the predictability of upcoming material in the source language input, lower surprisal leading to shorter translation latency and vice versa. Here, it would also be interesting to see how the effects of surprisal and relative frequency interact.

Acknowledgments

The research reported on has been funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102. We are grateful to the anonymous reviewers for their valuable suggestions on how to improve this paper.

References

Amos, Rhona M. & Pickering, Martin J. 2020. A theory of prediction in simultaneous interpreting. *Bilingualism: Language and Cognition* 23: 706–15. doi: 10.1017/S1366728919000671.

Amos, Rhona M. & Hartsuiker, Robert J. & Seeber, Kilian G. & Pickering, Martin J. 2022. Prediction during simultaneous interpreting. Evidence from the visual-world paradigm. *Cognition* 220: 104987. doi: 10.1016/j.cognition.2021.104987.

Arnold, Jennifer E. & Maria Fagnano, & Tanenhaus & Michael K. 2003. Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research* 32: 25–36. doi: 10.1023/A:1021980931292.

Bartłomiejczyk, Magdalena & Gumul, Ewa. 2024. Disfluencies and directionality in simultaneous interpreting. A corpus study comparing into-B and into-A interpretations from the European Parliament. *Translation & Interpreting* 16(1): 38–54. doi: 10.12807/ti.116201.2024.a03.

Bates, Douglas & Maechler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–48. doi: 10.18637/jss.v067.i01.

Bell, Alan & Jurafsky, Daniel & Fosler-Lussier, Eric & Girand, Cynthia & Gregory, Michelle & Gildea, Daniel. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113(2): 1001–1024. doi: 10.1121/1.1534836.

Bell, Alan & Brenier, Jason M. & Gregory, Michelle & Girand, Cynthia & Jurafsky, Dan. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60: 92–111. doi: 10.1016/j.jml.2008.06.003.

Boye, Kasper & Harder, Peter. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88(1): 1–44. doi: 10.1353/lan.2012.0020.

Brysbaert, Marc & Buchmeier, Matthias & Conrad, Markus & Jacobs, Arthur M & Bölte, Jens & Böhl, Andrea. 2011. The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental psychology* 58(5): 412–24. doi: 10.1027/1618-3169/a000123.

Brysbaert, Marc & Mandera, Paweł & Keuleers, Emmanuel. 2017. The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science* 27(1): 45–50. doi: 10.1177/0963721417727521.

Chmiel, Agnieszka, 2021. Effects of simultaneous interpreting experience and training on anticipation, as measured by word-translation latencies. *Interpreting* 23(1): 18–44. doi: 10.1075/intp.00048.chm.

Chmiel, Agnieszka & Koržinek, Danijel & Kajzer-Wietrzny, Marta & Janikowski, Przemysław & Jakubowski, Przemysław, & Polakowska, Dominika. 2022. Fluency parameters in the Polish Interpreting Corpus (PINC). In Kajzer-Wietrzny, Marta & Ferraresi, Adriano & Ivaska, Ilmari & Bernardini, Silvia (eds.), *Mediated discourse at the European Parliamen. Empirical investigations* (volume 19). Language Science Press. 63–91. doi: 10.5281/zenodo.6977042

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* 32(5): 905–23. doi: 10.1080/0907676X.2023.2218553.

Chmiel, Agnieszka & Kajzer-Wietrzny, Marta & Koržinek, Danijel & Jakubowski, Dariusz & Janikowski, Przemysław. 2024. Syntax, stress, and cognitive load. On syntactic processing in simultaneous interpreting. *Translation, Cognition & Behavior* 7(1): 22–47. doi: 10.1075/tcb.00091.chm.

Clark, Herbert H. & Fox Tree, Jean E. 2002. Using uh and um in spontaneous speaking. *Cognition* 84(1): 73–111. doi: 10.1016/S0010-0277(02)00017-3.

Corley, Martin & Stewart, Oliver W. 2008. Hesitation disfluencies in spontaneous speech. The meaning of um. *Language and Linguistics Compass* 2: 589–602. doi: 10.1111/j.1749-818X.2008.00068.x.

Crocker, Matthew W. & Demberg, Vera & Teich, Elke. 2015. Information density and linguistic encoding (IdeaL). *KI – Künstliche Intelligenz* 30: 77–81. doi: 10.1007/s13218-015-0391-y.

Dammalapati, Samvit & Rajkumar, Rajakrishnan & Agarwal, Sumeet. 2019. Expectation and locality effects in the prediction of disfluent fillers and repairs in English speech. In *Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Student research workshop.* Minneapolis, Minnesota. Association for Computational Linguistics. 103–09. doi: 10.18653/v1/N19-3015.

Dammalapati, Samvit & Rajkumar, Rajakrishnan & Ranjan, Sidharth & Agarwal, Sumeet. 2021. Effects of duration, locality, and surprisal in speech disfluency prediction in English spontaneous speech. In *Proceedings of the Society for Computation in Linguistics 2021*. Held online February 14–19, 2021. 91–101. doi: 10.7275/s5yn-7d90.

Defrancq, Bart & Plevoets, Koen. 2018. Over-uh-load, filled pauses in compounds as a signal of cognitive load. In Russo, Mariachiara & Bendazzoli, Claudio & Defrancq, Bart (eds.), *Making way in corpus-based interpreting studies*. Singapore: Springer. 43–64. doi: 10.1007/978-981-10-6199-8_3.

Demberg, Vera & Keller, Frank. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2): 193–210. doi: 10.1016/j.cognition.2008.07.008.

Fox, John & Weisberg, Sandford. 2019. *An R companion to applied regression* (3rd edition). Thousand Oaks, CA: Sage.

Gelman, Andrew & Stern, Hal. 2006. The difference between 'significant' and 'not significant' is not itself statistically significant. *The American Statistician* 60(4): 328–31. doi: 10.1198/000313006X152649.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics. Experiments in spontaneous speech*. London: Academic Press.

Wickham, Hadley. 2016. *ggplot2. Elegant graphics for data analysis*. New York: Springer.

Hauk, Olaf & Pulvermüller, Friedmann. 2004. Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology* 115(5): 1090–103. doi: 10.1016/j.clinph.2003.12.020.

Heller, Daphna & Arnold, Jennifer E. & Klein, Natalie & Tanenhaus, Michael K. 2015. Inferring difficulty. Flexibility in the real-time processing of disfluency. *Language and Speech* 58(2): 190–203. doi: 10.1177/0023830914528107.

Van Heuven, Walter J B & Mandera, Pawel & Keuleers, Emmanuel & Brysbaert, Marc. 2014. SUBTLEX-UK. A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology* 67(6): 1176–90. doi: 10.1080/17470218.2013.850521.

Hodzik, Ena & Williams, John N. 2017. Predictive processes during simultaneous interpreting from German into English. *Interpreting* 19(1):1–20. doi: 10.1075/intp.19.1.01hod.

Huettig, Falk & Guerra, Ernesto. 2019. Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research* 1706: 196–208. doi: 10.1016/j.brainres.2018.11.013.

Huettig, Falk & Audring, Jenny & Jackendoff, Ray. 2022. A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition* 224: 105050. doi: 10.1016/j.cognition.2022.105050.

Jiang, Xinlei & Jiang, Yue. 2020. Effect of dependency distance of source text on disfluencies in interpreting. *Lingua* 243: 102873. doi: 10.1016/j.lingua.2020.102873.

Kajzer-Wietrzny, Marta & Ivaska, Ilmari & Ferraresi, Adriano. 2024. Fluency in rendering numbers in simultaneous interpreting. *Interpreting* 26(1): 1–23. doi: 10.1075/intp.00092.kaj.

Kosmala, Loulou & Crible, Ludivine. 2022. The dual status of filled pauses. Evidence from genre, proficiency and co-occurrence. *Language and Speech* 65(1): 216–39. doi: 10.1177/00238309211010862.

Kuperberg, Gina R. & Jaeger, T. Florian. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31(1): 32–59. doi: 10.1080/23273798.2015.1102299.

Kuperman, Victor & Schroeder, Sascha & Gnetov, Daniil. 2024. Word length and frequency effects on text reading are highly similar in 12 alphabetic languages. *Journal of Memory and Language* 135: 104497. doi: 10.1016/j.jml.2023.104497.

Kutas, Marta & DeLong, Katherine A. & Smith, Nathaniel J. 2011. A look around at what lies ahead. Prediction and predictability in language processing. In Bar, Mosche (ed.), *Predictions in the brain. Using our past to generate a future.* New York: Oxford University Press. 190–207. doi: 10.1093/acprof:oso/9780195395518.003.0065.

Lange, Violaine Michel & Messerschmidt, Maria & Harder, Peter & Siebner, Hartwig Roman & Boye, Kasper. 2017. Planning and production of grammatical and lexical verbs in multi-word messages. *PLoS ONE* 12(11): e0186685. doi: 10.1371/journal.pone.0186685.

Lapshinova-Koltunski, Ekaterina & Pollkläsener, Christina & Przybyl, Heike. 2022. Exploring explicitation and implicitation in parallel interpreting and translation corpora. *The Prague Bulletin of Mathematical Linguistics* 119: 5–22. doi: 10.14712/00326585.020.

Liu, Yiguang & Hintz, Florian & Liang, Junying & Huettig, Falk. 2022. Prediction in challenging situations. Most bilinguals can predict upcoming semantically-related words in their L1 source language when interpreting. *Bilingualism. Language and Cognition* 25: 801–15. doi: 10.1017/S1366728922000232.

Malisz, Zofia & Brandt, Erika & Möbius, Bernd & Oh Yoon Mi & Andreeva, Bistra. 2018. Dimensions of segmental variability. Interaction of prosody and surprisal in six languages. *Frontiers in Communication / Language Sciences* 3(25): 1–18. doi: 10.3389/fcomm.2018.00025.

Montgomery, Douglas & Peck, Elizabeth A. 1992. *Introduction to linear regression analysis.* New York: Wiley.

Oh, Byung-Doh & Schuler, William. 2021. Why does surprisal from larger transformer-based language models provide a poorer fit to human reading times? *Transactions of the Association for Computational Linguistics* 11: 336–50. doi: 10.48550/arXiv.2212.12131.

R Core Team. 2024. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna.

Shain, Cory. 2024. Word frequency and predictability dissociate in naturalistic reading. *Open Mind. Discoveries in Cognitive Science* 8: 177–201. doi: 10.1162/opmi_a_00119.

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): 202–24. doi: 10.1075/tis.11.2.04ple.

Plevoets, Koen & Defrancq, Bart. 2018. The cognitive load of interpreters in the European Parliament. A corpus-based study of predictors for the disfluency uh(m). *Interpreting* 20(1): 1–28. doi: 10.1075/intp.00001.ple.

Przybyl, Heike & Lapshinova-Koltunski, Ekaterina & Menzel, Katrin & Fischer, Stefan & Teich, Elke. 2022a. EPIC UdS – creation and applications of a simultaneous interpreting corpus. *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*. Marseille, France: ELDA. 1193–1200.

Przybyl, Heike & Karakanta, Alina & Menzel, Katrin & Teich, Elke. 2022b. Exploring linguistic variation in mediated discourse: translation vs. interpreting. In Kajzer-Wietrzny, Marta & Bernardini, Silvia & Ferraresi, Adriano & Ivaska, Ilmari (eds.), *Mediated discourse at the European Parliament. Empirical investigations.* Berlin: Language Science Press. 191–218. doi: 10.5281/zenodo.6977050.

Qi, Peng & Zhang, Yuhao & Zhang, Yuhui & Bolton, Jason & Manning, Christopher D. 2020. Stanza. A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Online: ACL. 101–08. doi: 10.18653/v1/2020.acl-demos.14.

Radford, Alec & Wu, Jeffrey & Child, Rewon & Luan, David & Amodei, Dario & Sutskever, Ilya. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1(8): 9.

Schneider, Ulrike. 2014. Frequency, chunks and hesitations: A usage-based analysis of chunking in English. Freiburg i. Breisgau: Albert-Ludwigs-Universität. (Doctoral dissertation.). doi: 10.6094/978-3-928969-57-4.

Schweter, Stefan. 2020. *German GPT-2 model.* Zenodo.

Seifart, Frank & Strunk, Jan & Danielsen, Swintha & Hartmann, Iren & Pakendorf, Brigitte & Wichmann, Søren & Witzlack-Makarevich, Alena & de Jonge, Nivja H. & Bickel, Balthasar. 2018. Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences (PNAS)* 115(22): 5720–25. doi: 10.1073/pnas.1800708115

Shannon, Claude Elwood. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x.

Shlesinger, Miriam & Ordan, Noam. 2012. More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target. International Journal of Translation Studies*: 24(1): 43–60. doi: 10.1075/target.24.1.04shl.

Slaats, Sophie & Martin, Andrea. E. 2025. What's surprising about surprisal. *Computational Brain and Behavior* 8: 233–48. doi: 10.1007/s42113-025-00237-9.

Smith, Nathaniel J & Levy, Roger. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128(3): 302–19. doi: 10.1016/j.cognition.2013.02.013.

Teich, Elke & Martínez Martínez, José & Karakanta, Alina. 2020. Translation, information theory and cognition. In Alves, Fabio. and Jakobsen, Arnt L. (eds.), *The Routledge handbook of translation and cognition*. London: Routledge. 360–75.

Wang, Binhua & Li, Tao. 2014. An empirical study of pauses in Chinese-English simultaneous interpreting. *Perspectives* 23(1): 124–42. doi: 10.1080/0907676X.2014.948885.

Winter, Bodo. 2019. *Statistics for linguists. An introduction using R*. New York: Routledge.

Xu, Han & Liu, Kanglong. 2023. Syntactic simplification in interpreted English. Dependency distance and direction measures. *Lingua* 294: 103607. doi: 10.1016/j.lingua.2023.103607.

Zámečník, Jiří. 2019. Disfluency prediction in natural spoken language. Freiburg i. Breisgau: Albert-Ludwigs-Universität. (Doctoral dissertation.). doi: 10.6094/978-3-928969-84-0.

Christina Pollkläsener
University of Hildesheim
Dekanat FB 3
Universitätsplatz 1
31141 Hildesheim
Germany
e-mail: pollklaesener@uni-hildesheim.de

Maria Kunilovskaya
Saarland University
Campus
66123 Saarbrücken
Germany
e-mail: maria.kunilovskaya@uni-saarland.de

Elke Teich
Saarland University
Campus
66123 Saarbrücken
Germany
e-mail: e.teich@mx.uni-saarland.de