Um… Are Longer Filler Particles Predicted by Higher Surprisal?

Christina Pollkläsener (christina.pollklaesener@uni-saarland.de)[1]

[1]Department of Language Science and Technology, Saarland University

Filler particles (FPs), i.e. hesitation markers like *uh* or *um*, are often linked to problems in language planning of upcoming material. More recent research has refined this perspective by linking the occurrence of FPs to probabilistic properties of language: speakers are more likely to produce FPs before words that are unexpected or less predictable in context, as quantified by surprisal [1], [2]. Importantly, this has also been reported for simultaneous interpreting [3].

Beyond their mere presence, the phonetic form and duration of FPs may provide further insight into the extent of the underlying processing difficulty. Germanic languages like German and English usually distinguish between two major types of FPs: vowel (*uh*, *äh*) and vowel-nasal combinations (*um*, *ähm*). [4] proposed that speakers systematically choose between these FP forms to announce the initiation of either a minor or a major delay. They suggested that minor delays were associated with vocalic FPs and major delays with vowel-nasal combinations.

The present study explores the following two research questions: Do form and duration of FPs vary as a function of upcoming processing difficulty, operationalised as surprisal of the subsequent linguistic material? Does this effect differ between original, non-mediated speech and simultaneous interpreting?

To investigate this, we conduct a corpus study, analysing naturally occurring speech data. We use the German and English components of the EPIC-UdS Corpus [5], which contains audio recordings and transcripts of English and German speeches from Members of the European Parliament and their simultaneous interpretations into English and German respectively. In the corpus, each word is annotated with surprisal, which was estimated with GPT-2 language models (English: [9]; German: [10]). Filler particles and other disfluencies were excluded when estimating surprisal. The transcripts are time-aligned with the audio using WebMAUS [6]. For FP annotation, we follow an adapted scheme based on [7] (cf. Figure 1). Annotations are carried out in Praat [8]. The first tier (ORT-MAU) contains the time-aligned transcript by WebMAUS. Two additional tiers are introduced. In the FP-seg tier, pauses (*s*) are annotated separate from FPs. In the FP-full tier, FPs are grouped with their surrounding pause context. Pause context is encoded using plus and minus symbols (e.g., *+uh−*), where minus indicates adjacent silent intervals and plus indicates continuous speech. When several FPs appear in a row (e.g. *uh um*, *uh uh uh*), they are annotated separately in the FP-seg tier and marked as *cluster* in the FP-full tier.

For analyses, we run several mixed-effects regression models. Our response variables are FP duration (once for each individual FP as in FP-seg and once including pauses and clusters as in FP-full) and FP type (*uh* vs. *um* vs. *cluster*). Our main predictor is the surprisal of the next word following the FP. We also fit a categorical variable for text type (original vs. interpreted) and interactions between surprisal and text type to investigate possible differences in the effect of surprisal on interpreting and non-mediated speech. All models include lexical frequency of the subsequent word as a control predictor. We fit random intercepts for speaker and individual speech. Results of this study will be presented at the poster.

References

[1] Samvit Dammalapati, Rajakrishnan Rajkumar, and Sumeet Agarwal. Effects of duration, locality, and surprisal in speech disfluency prediction in english spontaneous speech. *Proceedings of the Society for Computation in Linguistics 2021*, 2021.

[2] Jiří Zámečník. *Disfluency prediction in natural spoken language*. Diss. Dissertation, Universität Freiburg, 2019.

[3] Christina Pollkläsener, Maria Kunilovskaya, Elke Teich. Surprisal explains the occurrence of filler particles in simultaneous interpreting. *SKASE Journal of Translation and Interpretation*, 18(2): 55–78, 2025.

[4] Herbert H. Clark, Jean E. Fox Tree. Using uh and um in spontaneous speaking. *Cognition* 84(1): 73-111, 2002.

[5] Maria Kunilovskaya, Christina Pollkläsener. EPIC-EuroParl-UdS: Information-Theoretic Perspectives on Translation and Interpreting. *Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026)*, European Language Resources Association (ELRA), 6998–7013, 2026.

[6] Florian Schiel. A Statistical Model for Predicting Pronunciation. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, Paper 195, 2015.

[7] Beeke Muhlack, Jürgen Trouvain, Michael Jessen. Distributional and Acoustic Characteristics of Filler Particles in German with Consideration of Forensic-Phonetic Aspects. *Languages* 8(2), 2023.

[8] Paul Boersma, David Weenink. *Praat: Doing Phonetics by Computer*. Computer program, Version 6.4.48, 2025.

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

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

Figures and Tables

Figure 1: Example of FP annotation in praat.