Segmentation and Annotation of Interpreting Units for Semantic Transfer Analysis

The present paper focuses on methodological questions related to the analysis of HeiCIC (Heidelberg Conference Interpreting Corpus), a corpus of simultaneously interpreted speeches by interpreting trainees and professional interpreters. We are concerned with identifying a unit of interpreting that allows for quantitative and qualitative analyses and captures changes in semantic transfer between source text (ST) and target text (TT) segments as a function of cognitive load.

In interpreting studies, different identifying criteria for appropriate units have been proposed, drawing on research into spoken language as well as translation (Fiehler et al. 2004, Grupo Val.Es.Co 2014, Alves et al. 2020). While researchers agree that functional units based on semantic criteria represent most accurately the units processed by interpreters (Setton 1999, Pöchhacker 2016), their identification poses challenges due to their subjective nature. Other approaches propose surface-level indicators based on e.g. propositions (Goldman-Eisler, 1972, Dillinger 1994), clauses (Wehrmeyer 2020) or prosodic identifiers but do not account for simultaneity of cognitive processes and the use of interpreting strategies. To our knowledge, no SI corpora exist with comprehensive segmentation and alignment below the sentence level. Current research into SI considers word- or sentence-level features or individual phenomena, as applied e.g. in EuroParl, EPIC and EPTIC (Bernardini et al. 2016, Dayter 2021, Gumul 2021, Lapshinova-Koltunski et al. 2022, Plevoets and Defrancq 2021). While some of these features may highlight individual traits of cognitive load or ST and TT relations, they do not represent the magnitude of effects or relate features to types of cognitive processing.

The English-German subcorpus of HeiCIC in focus here contains transcripts in both directions and several interpretations of the same original (currently ca. 117h, 636.400 tokens). Segmentation and alignment is combined with multilayer annotation including automatic analysis (tokenization, POS tagging), semi-automatic extraction of problem triggers and manual feature annotation. Our current research objective is to investigate fine-grained variation in types of semantic transfer (e.g. subtypes of explicitation and implicitation) as a function of cognitive load (Kunz et al. 2021). We cross-reference these results with interpreters' preparation strategies and their level of expertise. Our notion of interpreting units brings together information chunks in the ST and TT and provides the basis for manual segmentation and alignment as well as semantic transfer analysis of the whole corpus. We consider interpreting units as self-contained units of information which can potentially be completely processed. For ST segmentation, we use semantic and syntactic criteria below the sentence and clause boundaries to trace cognitive efforts in SI related to speech comprehension and short-term memory capacity, distinguishing between segments that consist of a clause with all required constituents for syntactic completeness and segments that constitute optional additions. ST units are aligned with TT units based on semantic indicators, enabling a comparative analysis of structural and semantic changes and identification of production effects.

The greatest challenges for segmentation and alignment, which also inhibit automatic processing, are incomplete structures on different linguistic levels. These however may be related to language contrast, directionality or spoken language, or be indicative of cognitive processes of SI. For instance, we may capture differences in incomplete structures between interpreting outputs of trainees and those of professionals which are due to varying degrees of cognitive load and use of different types of interpreting strategies (Kalina 1998). Apart from our own research, our approach will permit research into other areas of interest and may serve to identify patterns for automatic extraction and analysis of parallel interpreting corpora.

References

Alves, F., A. Pagano, S. Neumann, E. Steiner and S. Hansen-Schirra (2010). Translation units and grammatical shifts. Translation and Cognition. G. M. Shreve and E. Angelone**:** 109-142.

Bernardini, S., A. Ferraresi and M. Miličević (2016). "From EPIC to EPTIC — Exploring simplification in interpreting and translation from an intermodal perspective." Target. International Journal of Translation Studies 28(1): 61-86.

Dayter, D. (2021). "Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological richness, and position in speech event." Meta 65(3): 594-617.

Dillinger, M. (1994). Comprehension during interpreting. What do interpreters know that bilinguals don't? Bridging the Gap: Empirical research in simultaneous interpretation. S. Lambert and B. Moser-Mercer. Amsterdam, Benjamins**:** 155-190.

Fiehler, R., B. Barden, M. Elstermann, B. Kraft, B. Barden, M. Elstermann and B. Kraft, Eds. (2004). Eigenschaften gesprochener Sprache. Studien zur deutschen Sprache. Tübingen, Narr.

Goldman-Eisler, F. (1972). "Segmentation of Input in Simultaneous Translation." Journal of psycholinguistic research 1: 127-140.

Gumul, E. (2021). "Explicitation and cognitive load in simultaneous interpreting." Interpreting. International Journal of Research and Practice in Interpreting 23 (1): 45-75.

Kalina, S. (1998). Strategische Prozesse beim Dolmetschen: theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen. Tübingen, Narr.

Kunz, K., Stoll, Ch. and Klüber, E. (2021). HeiCiC: A simultaneous interpreting corpus combining product and pre-process data. In Proceedings for the First Workshop on Modelling Translation: Translatology in the Digital Age, pages 8–14, online

Lapshinova-Koltunski, E., C. Pollkläsener and H. Przybyl (2022). "Exploring Explicitation and Implicitation in Parallel Interpreting and Translation Corpora." The Prague Bulletin of Mathematical Linguistics 119: 5-22.

Plevoets, K. and B. Defrancq (2021). Imported load in simultaneous interpreting. Multilingual Mediated Communication and Cognition. R. Muñoz Martín and S. L. Halverson. London, Routledge**:** 18-43.

Pöchhacker, F. (2016). Segmentation. Routledge Encyclopedia of Interpreting Studies. F. Pöchhacker. London, Routledge**:** 367-368.

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

Val.Es.Co., G. (2014). "Las unidades del discurso oral : la propuesta Val.Es.Co. de segmentación de la conversación (coloquial)." Estudios de lingüística del español 35: 11-71.

Wehrmeyer, E. (2020). "Shifts in signed media interpreting." International Journal of Corpus Linguistics 25: 270-296.