

Eva Klüber/Kerstin Kunz

Segmentation and Annotation of Interpreting Units for Semantic Transfer Analysis

Keywords: cognitive load; segmentation; annotation; simultaneous interpreting

The present paper focuses on methodological questions related to the analysis of HeiCIC (Heidelberg Conference Interpreting Corpus), a corpus of simultaneously interpreted (SI) 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 units that most accurately represent a chunk of speech processed as one have been proposed, drawing on research into spoken language as well as translation (Fehler et al. 2004; Grupo Val.Es.Co 2014; Alves et al. 2020). While researchers agree that functional units based on semantic criteria resemble most closely 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/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 are 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 (IU) brings together information chunks in the ST and TT and provides the basis for manual segmentation and alignment. To allow for a transfer analysis of the whole corpus that also yields information about cognitive requirements of simultaneous interpreting, IUs are defined on the basis of structural, semantic and functional/processing criteria. We consider IUs as self-contained units of information which can potentially

be completely processed. For ST segmentation we make use of syntactic dependencies below the sentence and clause boundaries to identify interdependent elements. We further analyse the semantic content of an ST segment to determine whether it fulfils the minimum criteria of an informative and independent unit whose understanding does not require further additions. Not all identified segments can fulfil our criteria of independent units, due to spoken language features and language contrasts between ST and TT. We therefore distinguish between segments that consist of a clause with all required constituents for syntactic completeness, segments that lack syntactic completeness, and segments that constitute optional additions. Segmentation below the clause boundary also helps us to model cognitive requirements of SI such as ST comprehension and short-term memory capacity.

ST Segment 1	addition	Basically, when we are looking at a star,
ST Segment 2	main segment	the light from the star is passing through a lot of gas and dust
ST Segment 3	addition	in our universe [ehm]
ST Segment 4	main segment	and this gas [ehm] scatters the blue light [...]

Table 1: ST segmentation

ST units are aligned with TT units based on semantic indicators, so that structural transformations as well as semantic changes between ST and TT are revealed. This further allows to identify production efforts in the TT.

ST Seg 1	Basically, when we are looking at a star,	Wenn man sich einen Stern ansieht,	TT Seg 1
ST Seg 2	the light from the star is passing through a lot of gas and dust	dann wird das Licht durch Gas und Staub gebrochen.	TT Seg 2
ST Seg 3	in our universe [ehm]		
ST Seg 4	and this gas [ehm] scatters the blue light from the star,	Und das führt eben zu diesem blauen Licht,	TT Seg 3
		das wir sehen.	TT Seg 4

Table 2: TT alignment and segmentation

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. In the future, we plan to include an analysis of phonetic indicators, such as hesitation markers, to confirm processing units and analyse the development of interpreting units

in relation to experience by further examining data produced by interpreting students and professional interpreters.

References

- Alves, Fabio/Pagano, Adriana/Neumann, Stella/Steiner, Erich/Hansen-Schirra, Sylvia (2010): Translation units and grammatical shifts. In: Shreve, Gregory M./Angelone, Erik (eds.): *Translation and Cognition*, pp. 109–142.
- Bernardini, Silvia/Ferraresi, Adriano/Miličević, Maja (2016): From EPIC to EPTIC — Exploring simplification in interpreting and translation from an intermodal perspective. In: *Target. International Journal of Translation Studies* 28(1), pp. 61–86.
- Dayter, Daria (2021): Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological richness, and position in speech event. In: *Meta* 65(3), pp. 594–617.
- De Groot, Annette/Christoffels, Ingrid (2006): Language control in bilinguals: Monolingual tasks and simultaneous interpreting. In: *Bilingualism: Language and Cognition*, 9(2), pp. 189–201.
- Dillinger, Mike (1994): Comprehension during interpreting. What do interpreters know that bilinguals don't? In: Lambert, Sylvie/Moser-Mercer, Barbara (eds.): *Bridging the Gap: Empirical research in simultaneous interpretation*. Amsterdam, Benjamins, pp. 155–190.
- Fiehler, Reinhard/Barden, Birgit/Elstermann, Mechthild/Kraft, Barbara/Barden, Birgit/Elstermann, Mechthild (eds.) (2004): *Eigenschaften gesprochener Sprache. Studien zur deutschen Sprache*. Tübingen: Narr.
- Goldman-Eisler, Frieda (1972): Segmentation of Input in Simultaneous Translation. In: *Journal of psycholinguistic research* (1), pp. 127–140.
- Gumul, Ewa (2021): Explicitation and cognitive load in simultaneous interpreting. In: *Interpreting. International Journal of Research and Practice in Interpreting* 23(1), pp. 45–75.
- Kalina, Sylvia (1998): *Strategische Prozesse beim Dolmetschen: theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*. Tübingen: Narr.
- Kunz, Kerstin/Stoll, Christoph/Klüber, Eva (2021): HeiCiC: A simultaneous interpreting corpus combining product and pre-process data. In: *Proceedings for the First Workshop on Modelling Translation: Translationology in the Digital Age*, pp. 8–14.
- Lapshinova-Koltunski, Ekaterina/Pollkläsener, Christina/Przybyl, Heike (2022): Exploring Explicitation and Implication in Parallel Interpreting and Translation Corpora. In: *The Prague Bulletin of Mathematical Linguistics* (119), pp. 5–22.
- Plevoets, Koen/Defrancq, Bart (2021): Imported load in simultaneous interpreting. In: Muñoz Martín, Ricardo/Halverson Sandra L. (eds.): *Multilingual Mediated Communication and Cognition*. London: Routledge, pp. 18–43.
- Pöchhacker, Franz (2016): Segmentation. In: Pöchhacker, Franz (ed.): *Routledge Encyclopedia of Interpreting Studies*. London: Routledge, pp. 367–368.
- Setton, Robin (1999): *Simultaneous interpretation: a cognitive-pragmatic analysis*. Amsterdam: Benjamins.
- Val.Es.Co., Grupo (2014): Las unidades del discurso oral : la propuesta Val.Es.Co. de segmentación de la conversación (coloquial). In: *Estudios de lingüística del español* (35), pp. 11–71.
- Wehrmeyer, Ella (2020): Shifts in signed media interpreting. In: *International Journal of Corpus Linguistics* (25), pp. 270–296.

Contact information

Eva Klüber

Institut für Übersetzen und Dolmetschen, Universität Heidelberg

ICLC-10 2023 – Segmentation and Annotation of Interpreting Units for Semantic Transfer Analysis

eva.klueber@iued.uni-heidelberg.de

Kerstin Kunz

Institut für Übersetzen und Dolmetschen, Universität Heidelberg

kerstin.kunz@iued.uni-heidelberg.de