

Katrin Menzel

Initialisms in English and German European Parliament Data

Keywords: Translation corpora; interpreting corpora; shortening strategies; multi-word sequences; political discourse; English-German contrasts in word-formation

This paper addresses the question of how English (EN) and German (DE) use initialisms in European Parliament original and translated / interpreted discourse in the sentence-aligned Europarl_UdS (Karakanta et al., 2018) and EPIC-UdS (Przybyl et al., 2022) corpora. Europarl-UdS includes the European Parliament interventions in their officially published versions and their translations. EPIC-UdS with spoken data is smaller than Europarl-UdS and consists of transcripts of these European Parliament speeches and their simultaneous interpretation, without any corrections with respect to the spoken signal.

In general, initialisms as minimalistic submorphemic lexical units functioning as shortenings of multi-word sequences can be expected to be used rather similarly in the English and German data for items referring to international institutions, groups, projects and policies as proper nouns, e.g. *WTO*, *NATO*. Differences between English and German involve English forms for technical and scientific terms borrowed into German (e.g. EN/DE: *BSE*), but usually not from German into English. EU-specific and highly frequent terms may have a German equivalent with an initialism for the full form as a multiword expression or a closed compound (e.g. EN: *European Regional Development Fund / ERDF* – DE: *Europäischer Fonds für regionale Entwicklung / EFRE*, EN: *Gross domestic product / GDP* – DE: *Bruttoinlandsprodukt / BIP*). Some initialisms happen to be the same in both languages if the source expressions are structurally similar and involve cognates (e.g. *European Stability Mechanism / Europäischer Stabilitätsmechanismus – ESM*).

Initialisms can be demanding for interpreters in both English and German, and there are probably similar interpreting procedures for English and German initialisms that contribute to general differences between interpreted and translated texts. Initialisms in the original texts may be among the less expected textual items and have high surprisal values despite the general tendency of short codes to represent messages of high probability (Shannon 1948: 395). For interpreters, they might be potentially ambiguous or difficult to associate with their underlying full forms. This might lead, for instance, to disfluencies or a loss of lexical information or specificity in the interpreted texts or to the usage of units in untypical contexts with an even higher degree of unexpectedness for the recipients of the target texts.

The Europarl_UdS and EPIC-UdS data are queried and analysed by using CQPWeb to compare frequencies in the languages and production modes for initialisms, the contexts they are used in, their surprisal profiles and the respective translation / interpreting procedures. Various short forms look superficially similar, but

have been coined on the basis of different word formation processes (e.g. *EULEX* stands for *European Union Rule of Law Mission in Kosovo*). Therefore, one step involved the development of a categorisation/annotation scheme. The results of the analysis show, for instance, that German spoken and written originals use initialisms mainly for foreign terms, and often combine them with additional nouns to form compounds, e.g. EN: *'the debate on BSE'* -> DE: *'die BSE-Debatte'*, EN: *'at the WTO'* -> DE: *'auf WTO-Ebene'*. In spoken texts from both languages, initialisms are used very flexibly in word-formation processes, e.g. in adjective compounds such as *'WTO-compatible'* / *'WTO-kompatibel'*, which are less frequent in written EU texts. Interpreters sometimes opt for leaving out initialisms with high surprisal values entirely, probably due to cognitive overload. If interpreters opt for adding paraphrases or additional words, they tend to use general (and sometimes erroneous) terms, while in the written data, initialisms in both languages are more often retained adequately, paraphrased with specific vocabulary or spelt out as full form in the translations (e.g. DE: *'die EVP-Fraktion'* -> EN: *'the Group of the European People's Party'*). Interpreters introduce initialisms in certain contexts as impicitation strategies to save time, while written translations in this field tend to be at least as explicit as their source texts. For instance *'Interim-WPA mit dem Pazifischen Raum'* is used in a German interpretation for *'Interim Economic Partnership Agreement with the Pacific'*, while the written translation of the same English term in this context contains *'Interim-Wirtschaftspartnerschaftsabkommen mit Staaten im Pazifischen Ozean'*. The results show that surprisal values are typically higher for initialisms than for individual components of multiword terms and proper nouns.

In sum, initialisms as word-like shortening devices for multi-word sequences are important in both languages in these registers with expert-to-expert and expert-to-general public interaction. In English, an advantage of initialisms in written texts is to reduce the number of orthographic words in noun pile-ups for specialized vocabulary, while in German they reduce long closed compounds with technical meanings. In spoken language, they save time and establish a sense of shared knowledge and expertise among the speaker and the audience. In both written and spoken texts and in interpreted and translated texts in English and German, they are characterized by high information density, measured by surprisal on the basis of the previous words as predictors.

References

- Karakanta, Alina, Mihaela Vela and Elke Teich (2018): Europarl-UdS: Preserving metadata from parliamentary debates. In: Darja Fišer, Maria Eskevich & Franciska de Jong (eds.), Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018).
- Lapshinova-Koltunski, Ekaterina/Bizzoni, Yuri/Przybyl, Heike/Teich, Elke (2021): Found in translation/interpreting: Combining data-driven and supervised methods to analyse cross-linguistically mediated communication. In: Proceedings of the Workshop on Modelling Translation: Translatology in the Digital Age (MoTra21), Association for Computational Linguistics, pp. 82–90.

- Przybyl, Heike/Karakanta, Alina, Menzel, Katrin/Teich, Elke (2022): Exploring linguistic variation in mediated discourse: Translation vs. interpreting. In: Kajzer-Wietrzny, Marta/Ferraresi, Adriano/Ivaska, Ilmari/Bernardini, Silvia (eds.): Mediated discourse at the European Parliament: Empirical investigations. Berlin: Language Science Press, pp. 191–218.
- Shannon, Claude E. (1948): A mathematical theory of communication. In: Bell Systems Technical Journal 27, pp. 379–423.

Contact information

Katrin Menzel

Universität des Saarlandes

k.menzel@mx.uni-saarland.de