

Initialisms in English and German European Parliament Data

This paper addresses the question of how English (EN) and German (DE) use initialisms in European Parliament original and translated 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 consists of transcripts of these speeches and their simultaneous interpretation, without any corrections with respect to the spoken signal.

We expect that initialisms for multi-word sequences are used rather similarly in the English and German data for items referring to international institutions, groups, projects and policies, e.g. WTO, NATO, while domain-specific terms more often might have a German equivalent (e.g. EN: GDP – DE: BIP), but may also correspond to borrowed forms (e.g. EN/DE: BSE). Initialisms can be demanding for interpreters and we expect to see similar effects in the translation procedures for initialisms in both languages.

The data will be queried and analysed by using CQPWeb to compare the languages and production modes for initialisms, their integration into the syntactic contexts and their translation procedures. German uses initialisms mainly for foreign terms in these data, and often combines them with an additional noun to form a hyphenated compound, 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 even more flexibly in word-formation processes, e.g. in adjective compounds such as WTO-compatible / WTO-kompatibel. Particularly in the spoken data, interpreters sometimes opt for leaving out initialisms entirely or for retaining them in contexts where the listeners might not be familiar with them (e.g. EN: PCBs [polychlorinated biphenyls] -> DE: PCBs). They also replace some initialisms by more general (and sometimes erroneous) terms, while in the written translated data, initialisms in both languages are more often retained adequately or spelt out as full form (e.g. DE: die EVP-Fraktion -> EN: the Group of the European People's Party).

In sum, initialisms as word-like shortening devices for multi-word sequences are important in this type of political communication due to their aesthetic, expressive and mnemonic functions. In this registers with much expert-to-expert and expert-to-general public interaction, initialisms are often easier to use than full forms. Especially acronymic initialisms, i.e. combinations of initial letters read as words, can be easier to pronounce or memorize and to circulate at national and international levels than expanded expressions (Anonymous, submitted).

References:

Anonymous Submitted.

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, Yuri Bizzoni, Heike Przybyl and Elke Teich. 2021. Found in translation/interpreting: combining data-driven and supervised methods to analyse cross-linguistically mediated communication. Proceedings of the Workshop on Modelling Translation: Translatology in the Digital Age (MoTra21), Association for Computational Linguistics, pp. 82-90, online.

Przybyl, Heike, Alina Karakanta, Katrin Menzel, and Elke Teich. 2022. Exploring linguistic variation in mediated discourse: translation vs. interpreting. In M. Kajzer-Wietrzny, S. Bernardini, A. Ferraresi, and I. Ivaska (Eds.), *Empirical investigations into the forms of mediated discourse at the European Parliament*, Translation and Multilingual Natural Language Processing. Berlin: Language Science Press.