SwissSLi: the Multi-parallel Sign Language Corpus for Switzerland

Zifan Jiang et al.

In this work, we introduce SwissSLi, the first sign language corpus that contains parallel data of all three Swiss sign languages, namely Swiss German Sign Language (DSGS), French Sign Language of Switzerland (LSF-CH), and Italian Sign Language of Switzerland (LIS-CH). The data underlying this corpus originates from television programs in three spoken languages: German, French, and Italian. The programs have for the most part been translated into sign language by deaf translators, resulting in a unique, up to six-way multi-parallel dataset between spoken and sign languages. We describe and release the sign language videos and spoken language subtitles as well as the overall statistics and some derivatives of the raw material. These derived components include cropped videos, pose estimation, phrase/sign-segmented videos, and sentence-segmented subtitles, all of which facilitate downstream tasks such as sign language transcription (glossing) and machine translation. The corpus is publicly available on the SWISSUbase data platform for research purposes only under a CC BY-NC-SA 4.0 license