SignON Sign Language Translation: Progress and Challenges

Vincent Vandeghinste[†a], Dimitar Shterionov[*], Mirella De Sisto[*], Aoife Brady[‡], Mathieu De Coster[§]
Lorraine Leeson[¶], Josep Blat[**], Frankie Picron[††], Marcello Paolo Scipioni[‡‡]
Aditya Parikh[§§], Louis ten Bosch[§§], John O'Flaherty[‖], Joni Dambre[§], Jorn Rijckaert[x]
Bram Vanroy[a], Victor Ubieto Nogales[**], Santiago Egea Gomez[**], Ineke Schuurman[a]
Gorka Labaka[b], Adrián Núñez-Marcos[b], Irene Murtagh[c],Euan McGill[**], Horacio Saggion[**]
[*]Tilburg University, [†]Instituut voor de Nederlandse Taal, [‡]ADAPT, [§]Ghent University,
[¶]Trinity College Dublin,[**]Universitat Pompeu Fabra,[††]European Union of the Deaf,
[‡‡]Fincons,[§§]Radboud University,[‖]mac.ie,[x]Vlaams Gebarentaalcentrum, [a]KU Leuven,
[b]University of the Basque Country UPV/EHU, [c]TU Dublin

SignON[1] is a Horizon 2020[2] project, running from January 2021 until December 2023, addressing the lack of technology and services for MT between sign languages (SLs) and spoken languages (SpLs), through an inclusive, human-centric solution, contributing to the repertoire of communication media for deaf, hard of hearing (DHH) and hearing individuals. Even though there are estimates that over 70 million DHH individuals have SLs as their primary means of communication, SLs are often not targeted by new language technologies, due to challenges, such as the scarcity of data and the lack of a standardized written representation. This paper presents an update of the project status, describing how we address the challenges and peculiarities of SLMT.

We built an MT framework between SLs and SpLs, in all possible combinations, focusing on Irish, Dutch, Flemish, Spanish and British SL and on Irish, Dutch, Spanish and English SpLs (spoken and written). To limit the computational complexity and allow the effective development of components in parallel, we develop a translation pipeline that employs an interlingual representation (InterL) (Figure 1). Inputs can be an SpL utterance in audio or text or an SL utterance in video. The input is processed via the corresponding component: automatic speech recognition (ASR) converts audio into text; SL recognition (SLR) converts SL videos into latent representations. The integration of all of these components is currently ongoing. We develop ASR for both typical and atypical speech, such as speech of DHH persons. A use case sub-project collects speech data from this specific user group. Both conventional 'modular' approaches as well as more recently developed end-to-end approaches based on deep learning (DL) are employed.

SLR uses a pose estimator (Lugaresi et al., 2019) and post-processing of the predicted keypoints. This yields robust representations: missing data are imputed and keypoints are normalised to account for camera position. These representations are further processed into embeddings, which are fine-tuned on SL data, using glosses as target labels. However, we do not predict glosses but extract visual embeddings which are used as input for the SL MT models.

We use mBART (Liu et al., 2020) for text-to-text translation, fine-tuned to also support Irish and SL-to-text translation, trained to work with visual embeddings coming from SLR. We also operationalise knowledge-based approaches. We use Abstract Meaning Representation (AMR) (Banarescu et al., 2013) as an InterL to "extract" meaning. mBART was fine-tuned on automatically translated versions of the AMR Bank 3.0 (Knight et al., 2020) to create a multilingual text-to-AMR model.[3] Because of the lack of SL data we work on a knowledge-based alternative and use rule-based methods for data-augmentation (Chiruzzo et al., 2022). Schuurman et al. (to appear) investigate whether SL WordNet ("SignNets") can be linked to existing WordNets or whether the difference in modality warrants its own approach.

The output of the InterL (AMR or embeddings) is decoded into the target language. In case of a target SL, this is a representation for avatar movement, such as BML (Behaviour Markup Lan-

[1]https://signon-project.eu/

[2]Research and Innovation Programme Grant Agreement No. 101017255

[3]https://huggingface.co/spaces/BramVanroy/text-to-amr

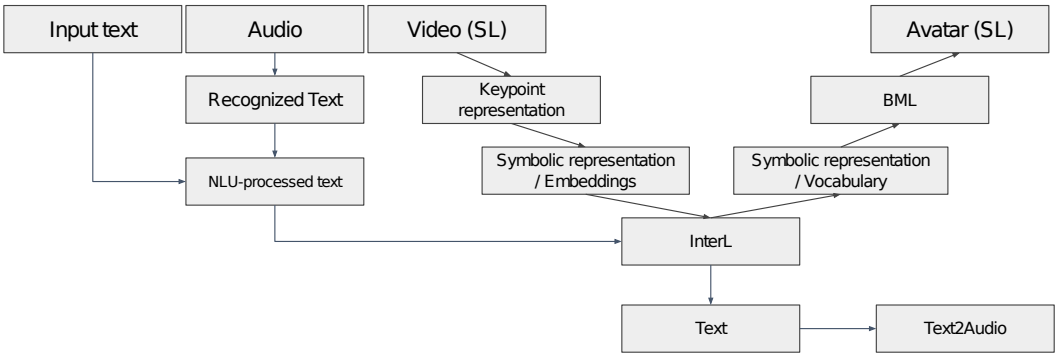

Figure 1: The SignON MT pipeline facilitating the translation between all supported sign and spoken languages.

guage) (Murtagh et al., 2022) or SiGML (Signing Gesture Markup Language). In case of SpLs it is text, which can be converted to speech through a text-to-speech system.

To allow users acces to the SignON services, we have developed a mobile app (for iOS and Android) that has access to the SignON MT pipeline.

Development of SLR and SLMT tools is slowed down due to resource scarcity and standardization issues in the available data. De Sisto et al. (2022) compare various SL corpora and machine learning datasets and propose a framework to unify the available resources and facilitate SL research. We have initiated a number of data collection efforts. Vandeghinste et al. (2022) compiled a corpus of Belgian COVID-19 press conferences, annotated with keypoints and speech recognition, providing a parallel VGT-NL dataset. GostParcSign (De Sisto et al., submitted) and NGT-HoReCo are two small datasets in which professional SL translators translate VGT into Dutch and Dutch into NGT, respectively. Another approach towards data collection is through the SignON ML app, which allows SL users to upload SL recordings and their associated translation in a written language.

SignON is in a continuous dialogue with target users. We regularly organize co-creation events (e.g. round tables, focus groups, and workshops) to receive feedback on the project's progress, which is then used to steer and refine further developments.

Conclusions Up till now we have conducted a significant amount of research in the fields of SLR, SL(M)T, SLS, ASR, (SL) linguistics, ethics, and others. We continue the development and testing of models as well as their validation by the community. We have co-developed the inference as well as ML Apps. We have established a fruitful co-creation that allows hearing, deaf and hard of hearing professionals and potential users to work together.

References

Banarescu, L., C. Bonial, S. Cai, et al. 2013. Abstract Meaning Representation for Sembanking. In *7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, August.

Chiruzzo, L., E. McGill, S. Egea-Gómez, and H. Saggion. 2022. Translating Spanish into Spanish Sign Language: Combining rules and data-driven approaches. In *LoResMT*.

De Sisto, M., V. Vandeghinste, S. Egea Gómez, et al. 2022. Challenges with Sign Language Datasets for Sign Language Recognition and Translation. In *LREC*, pages 2478–2487.

De Sisto, Mirella, Vincent Vandeghinste, and Dimitar Shterionov. submitted. GoSt-ParC-Sign: Gold Standard Parallel Corpus of Sign and spoken language. In *EAMT 2023*.

Knight, Kevin, Bianca Badarau, Laura Baranescu, et al. 2020. Abstract Meaning Representation (AMR) Annotation Release 3.0, January.

Liu, Yinhan, Jiatao Gu, Naman Goyal, et al. 2020. Multilingual Denoising Pre-training for Neural Machine Translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Lugaresi, C., J. Tang, H. Nash, et al. 2019. MediaPipe: A Framework for Perceiving and Processing Reality. In *Workshop at CVPR*.

Murtagh, Irene, Víctor Ubieto Nogales, and Josep Blat. 2022. Sign Language Machine Translation and the Sign Language Lexicon: A Linguistically Informed Approach. In *AMTA*, pages 240—-251.

Schuurman, I., T. Declerck, C. Brosens, et al. to appear. Are there just WordNets or also SignNets? In *Global WordNet Conference*.

Vandeghinste, V., B. Van Dyck, M. De Coster, et al. 2022. BeCoS Corpus: Belgian Covid-19 Sign Language Corpus. A Corpus for Training Sign Language Recognition and Translation. *CLIN Journal*, 12:7–17.