Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Sister Conferences Best Papers Track

Including Signed Languages in Natural Language Processing (Extended Abstract)*

Kayo Yin[1†], **Malihe Alikhani** [2‡]
[1]Carnegie Mellon University
[2]University of Pittsburgh
kayoy@cs.cmu.edu, malihe@pitt.edu

Abstract

Signed languages are the primary means of communication for many deaf and hard of hearing individuals. Since signed languages exhibit all the fundamental linguistic properties of natural language, we believe that tools and theories of Natural Language Processing (NLP) are crucial towards its modeling. However, existing research in Sign Language Processing (SLP) seldom attempt to explore and leverage the linguistic organization of signed languages. This position paper calls on the NLP community to include signed languages as a research area with high social and scientific impact. We first discuss the linguistic properties of signed languages to consider during their modeling. Then, we review the limitations of current SLP models and identify the open challenges to extend NLP to signed languages. Finally, we urge (1) the adoption of an efficient tokenization method; (2) the development of linguistically-informed models; (3) the collection of real-world signed language data; (4) the inclusion of local signed language communities as an active and leading voice in research.

1 Introduction

Natural Language Processing (NLP) has revolutionized the way people interact with technology through the rise of personal assistants and machine translation systems, to name a few. However, the vast majority of NLP models require a spoken language input (speech or text), thereby excluding around 200 different signed languages and up to 70 million deaf people[1] from modern language technologies.

Throughout history, Deaf communities fought for the right to learn and use signed languages, as well as for the recognition of signed languages as legitimate languages (§2). Indeed, signed languages are sophisticated communication modalities that are at least as capable as spoken languages in all manners, linguistic and social. However, in a predominantly oral society, deaf people are constantly encouraged to use spoken languages through lip-reading or text-based communication. The exclusion of signed languages from modern language technologies further suppresses signing in favor of spoken languages. This disregards the preferences of the Deaf communities who strongly prefer to communicate in signed languages both online and for in-person day-to-day interactions, among themselves and when interacting with spoken language communities. Thus, it is essential to make signed languages accessible.

To date, a large amount of research on Sign Language Processing (SLP) has been focused on the visual aspect of signed languages, led by the Computer Vision (CV) community, with little NLP involvement. This is not unreasonable, given that a decade ago, we lacked the adequate CV tools to process videos for further linguistic analyses. However, like spoken languages, signed languages are fully-fledged systems that exhibit all the fundamental characteristics of natural languages, and current SLP techniques fail to address or leverage the linguistic structure of signed languages (§3). This leads us to believe that NLP tools and theories are crucial to process signed languages. Given the recent advances in CV, this position paper argues that now is the time to incorporate linguistic insight into signed language modeling.

Signed languages introduce novel challenges for NLP due to their visual-gestural modality, simultaneity, spatial coherence, and lack of written form. By working on signed languages, the community will gain a more holistic perspective on natural languages through a better understanding of how meaning is conveyed by the visual modality and how language is grounded in visuospatial concepts.

Moreover, SLP is not only an intellectually appealing area but also an important research area with a strong potential to benefit signing communities. Examples of beneficial applications enabled by signed language technologies include better documentation of endangered sign languages; educational tools for sign language learners; tools for query and retrieval of information from signed language videos; personal assistants that react to signed languages; real-time automatic sign language interpretations, and more. Needless to say, in addressing this research area, researchers should work *alongside* and *under the direction of* deaf communities, and to the benefit of the signing communities' interest above all.

*The original paper was awarded *Best Theme Paper* at the 59th Annual Meeting of the Association for Computational Linguistics.

†Contact Author

‡Full author list: Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, Malihe Alikhani

[1]https://wfdeaf.org/our-work

After identifying the challenges and open problems to successfully include signed languages in NLP (§4), we emphasize the need to: (1) develop a standardized tokenization method of signed languages with minimal information loss for its modeling; (2) extend core NLP technologies to signed languages to create linguistically-informed models; (3) collect signed language data of sufficient size that accurately represents the real world; (4) involve and collaborate with the Deaf communities at every step of research. We refer readers to [Yin *et al.*, 2021b] for the full paper.

2 History of Signed Languages and Deaf Culture

Over the course of modern history, spoken languages were dominant so much so that signed languages struggled to be recognized as languages in their own right and educators developed misconceptions that signed language acquisition may hinder the development of speech skills. For example, in 1880, a large international conference of deaf educators called the *Second International Congress on Education of the Deaf* banned teaching signed languages, favoring speech therapy instead. It was not until the seminal work on American Sign Language (ASL) by [Stokoe, 1960] that signed languages started gaining recognition as natural, independent, and well-defined languages, which then inspired other researchers to further explore signed languages.

Nevertheless, antiquated notions that deprioritized signed languages continue to do harm and subjects many to linguistic neglect. Several studies have shown that deaf children raised solely with spoken languages do not gain enough access to a first language during their critical period of language acquisition. This language deprivation can lead to life-long consequences on the cognitive, linguistic, socioemotional, and academic development of the deaf.

Signed languages are the primary languages of communication for the Deafand are at the heart of Deaf communities. Failing to recognize signed languages as fully-fledged natural language systems in their own right has had harmful effects in the past, and in an increasingly digitized world, the NLP community has an important responsibility to include signed languages in its research. NLP research should strive to enable a world in which all people, including the Deaf, have access to languages that fit their lived experience.

3 Current State of SLP

In this section, we present the existing methods, resources, and tasks in SLP, and discuss their limitations to lay the ground for future research.

3.1 Existing Sign Language Resources

Now, we introduce the different formats of resources and discuss how they can be used for signed language modeling.

Bilingual dictionaries for signed language [Fenlon *et al.*, 2015] map a spoken language word or short phrase to a signed language video. One notable dictionary is, SpreadTheSign[2] is a parallel dictionary containing around 23,000 words with up to 41 different spoken-signed language pairs and more than 500,000 videos in total. While dictionaries may help create lexical rules between languages, they do not demonstrate the grammar or the usage of signs in context.

Fingerspelling corpora usually consist of videos of words borrowed from spoken languages that are signed letter-by-letter. They can be synthetically created [Dreuw *et al.*, 2006] or mined from online resources [Shi *et al.*, 2018]. However, they only capture one aspect of signed languages.

Isolated sign corpora are collections of annotated single signs. They are synthesized or mined from online resources [Li *et al.*, 2020], and can be used for isolated sign language recognition or for contrastive analysis of minimal signing pairs [Imashev *et al.*, 2020]. However, like dictionaries, they do not describe relations between signs nor do they capture coarticulation during signing, and are often limited in vocabulary size (20-1000 signs)

Continuous sign corpora contain parallel sequences of signs and spoken language. Available continuous sign corpora are extremely limited, containing 4-6 orders of magnitude fewer sentence pairs than similar corpora for spoken language machine translation. Moreover, the largest continuous sign language corpus contain only 1,150 hours, and only 50 of them are publicly available [Hanke *et al.*, 2020]. These datasets are usually synthesized [Hanke *et al.*, 2020] or recorded in studio conditions [Camgöz *et al.*, 2018], which does not account for noise in real-life conditions. Moreover, some contain signed interpretations of spoken language, which may not accurately represent native signing.

Availability Unlike the vast amount and diversity of available spoken language resources that allow various applications, signed language resources are scarce and currently only support translation and production. Unfortunately, most of the signed language corpora discussed in the literature are either not available for use or available under heavy restrictions and licensing terms. Signed language data is especially challenging to anonymize due to the importance of facial and other physical features in signing videos, limiting its open distribution, and developing anonymization with minimal information loss, or accurate anonymous representations is a promising research problem.

3.2 Sign Language Processing Tasks

The CV community has mainly led the research on SLP so far to focus on processing the visual features in signed language videos. As a result, current SLP methods do not fully address the linguistic complexity of signed languages. We survey common SLP tasks and limitations of current methods by drawing on linguistic theories of signed languages.

Detection Sign language detection is the binary classification task to determine whether a signed language is being used or not in a given video frame. While recent detection models [Borg and Camilleri, 2019] achieve high performance, we lack well-annotated data that include interference and distractions with non-signing instances for proper evaluation. A similar task in spoken languages is voice activity

[2]https://www.spreadthesign.com/

detection (VAD) [Sohn *et al.*, 1999], the detection of when a human voice is used in an audio signal. However, as VAD methods often rely on speech-specific representations such as spectrograms, they are not always applicable to videos.

Identification Sign language identification classifies which signed language is being used in a given video automatically. Existing works utilize the distribution of phonemes [Gebre *et al.*, 2013] or activity maps in signing space [Monteiro *et al.*, 2016] to identify the signed language in videos. However, these methods only rely on low-level visual features, while signed languages have several distinctive features on a linguistic level, such as lexical or structural differences [Kimmelman, 2014] which have not been explored.

Segmentation Segmentation consists of detecting the frame boundaries for signs or phrases in videos to divide them into meaningful units. Current methods resort to segmenting units loosely mapped to signed language units [Bull *et al.*, 2020], and does not leverage reliable linguistic predictors of sentence boundaries such as prosody in signed languages (i.e. pauses, sign duration, facial expressions, eye apertures) [Ormel and Crasborn, 2012].

Recognition Sign language recognition (SLR) detects and label signs from a video, either on isolated [Sincan and Keles, 2020] or continuous [Camgöz *et al.*, 2018] signs. Though some previous works have referred to this as "sign language translation", recognition merely determines the associated label of each sign, without handling the syntax and morphology of the signed language [Padden, 1988] to create a spoken language output. Instead, SLR has often been used as an intermediate step during translation to produce glosses from signed language videos.

Translation Sign language translation (SLT) commonly refers to the translation of signed language to spoken language. Current methods either perform translation with glosses [Camgöz *et al.*, 2018; Yin and Read, 2020a; Yin and Read, 2020b; Moryossef *et al.*, 2021] or on pose estimations and sign articulators from videos [Camgöz *et al.*, 2020], but do not, for instance, handle spatial relations and grounding in discourse to resolve ambiguous referents.

Production Sign language production consists of producing signed language from spoken language and often use poses as an intermediate representation to overcome challenges in animation. To overcome the challenges in generating videos directly, most efforts use poses as an intermediate representation, with the goal of either using computer animation or pose-to-video models to perform video production. Earlier methods generate and concatenate isolated signs [Stoll *et al.*, 2020], while more recent methods [Saunders *et al.*, 2020] autoregressively decode a sequence of poses from an input text. Due to the lack of suitable automatic evaluation methods of generated signs, existing works resort to measuring back-translation quality, which cannot accurately capture the quality of the produced signs nor its usability in real-world settings. A better understanding of how distinctions in meaning are created in signed language may help develop a better evaluation method.

4 Towards Including Signed Languages in Natural Language Processing

The limitations in the design of current SLP models often stem from the lack of exploring the linguistic possibilities of signed languages. We therefore invite the NLP community to collaborate with the CV community, for their expertise in visual processing, and signing communities and sign linguists, for their expertise in signed languages and the lived experiences of signers, in researching SLP. We believe that first, developing known tasks in standard NLP pipelines for signed languages will help us better understand how to model them, as well as provide valuable tools for higher-level applications.

Although these tasks have been thoroughly researched for spoken languages, they pose interesting new challenges in a different modality. We also emphasize the need for real-world data to develop such methods, and a close collaboration with signing communities to have an accurate understanding of how language technologies can benefit signers, all the while respecting their ownership of signed languages.

4.1 Building NLP Pipelines

Although signed and spoken languages differ in modality, we argue that as both express the syntax, semantics, and pragmatics of natural languages, fundamental theories of NLP can and should be extended to signed languages. NLP applications often rely on low-level tools such as tokenizers and parsers, so we invite more research efforts on these core NLP tasks that often lay the foundation of other applications. We also discuss what considerations should be taken into account for their development to signed languages and raise open questions that should be addressed.

Tokenization The vast majority of NLP methods require a discrete input. To extend NLP technologies to signed languages, we must first and foremost be able to develop adequate tokenization tools that maps continuous signed language videos to a discrete, accurate representation with minimal information loss. While existing SLP systems and datasets often use glosses as discrete lexical units of signed phrases, this poses three significant problems: (1) linear, single-dimensional glosses cannot fully capture the spatial constructions of signed languages, which downgrades downstream performance [Yin and Read, 2020b]; (2) glosses are language-specific and requiring new glossing models for each language is impractical given the scarcity of resources; (3) glosses lack standard across corpora which limits data sharing and adds significant overhead in modeling. We thus urge the adoption of an *efficient*, *universal*, and *standardized* method for tokenization of signed languages.

Core NLP Tools We encourage developing core NLP tools for signed languages, such as syntactic analysis, POS taggers, named entity recognition and coreference resolution [Yin *et al.*, 2021a]. These systems are all valuable tools that often lay the foundation of higher-level NLP systems. We believe that developing known NLP pipelines for signed languages will help us better understand how to model them. Moreover, even though signed and spoken languages are expressed in different modalities, they both express the grammar of natural languages, and fundamental theories of NLP can be extended to signed languages. Although these tasks have been thoroughly studied for some spoken languages, in the visual modality of signed languages, they pose interesting new challenges which we discuss further in [Yin *et al.*, 2021b].

4.2 Collect Real-World Data

Data is essential to develop any of the core NLP tools previously described, and current efforts in SLP are often limited by the lack of adequate data. We discuss the considerations to keep in mind when building datasets, challenges of collecting such data, and directions to facilitate data collection.

What is Good Signed Language Data? For SLP models to be deployable, they must be developed using data that represents the real world accurately. What constitutes an ideal signed language dataset is an open question, we suggest the following requirements: (1) a broad domain; (2) sufficient data and vocabulary size; (3) real-world conditions; (4) naturally produced signs; (5) a diverse signer demographic; (6) native signers; and when applicable, (7) dense annotations.

Challenges of Data Collection Collecting and annotating signed data inline with the ideal requires more resources than speech or text data, taking up to 600 minutes per minute of an annotated signed language video [Hanke *et al.*, 2020]. Moreover, annotation usually requires a specific set of knowledge and skills, which makes recruiting or training qualified annotators challenging. Additionally, there is little existing signed language data in the wild that are open to use, especially from native signers that are not interpretations of speech. Therefore, data collection often requires significant efforts and costs of on-site recording as well.

Automating Annotation To collect more data that enables the development of deployable SLP models, one useful research direction is creating tools that can simplify or automate parts of the collection and annotation process. One of the largest bottleneck in obtaining more adequate signed language data is the amount of time and scarcity of experts required to perform annotation. Therefore, tools that perform automatic parsing, detection of frame boundaries, extraction of articulatory features, suggestions for lexical annotations, and allow parts of the annotation process to be crowdsourced to non-experts, to name a few, have a high potential to facilitate and accelerate the availability of good data.

4.3 Practice Deaf Collaboration

Finally, when working with signed languages, it is vital to keep in mind *who* this technology should benefit, and *what* they need. Researchers in SLP must honor that signed languages belong to the Deaf community and avoid exploiting their language as a commodity.

Solving Real Needs Many efforts in SLP have developed intrusive methods (e.g. requiring signers to wear special gloves), which are often rejected by signing communities and therefore have limited real-world value. Such efforts are often marketed to perform "sign language translation" when they, in fact, only identify fingerspelling or recognize a very limited set of isolated signs at best. These approaches oversimplify the rich grammar of signed languages, promote the misconception that signs are solely expressed through the hands, and are considered by the Deaf community as a manifestation of audism, where it is the signers who must make the extra effort to wear additional sensors to be understood by non-signers. In order to avoid such mistakes, we encourage close Deaf involvement throughout the research process to ensure that we direct our efforts towards applications that will be adopted by signers, and do not make false assumptions about signed languages or the needs of signing communities.

Building Collaboration Deaf collaborations and leadership are essential for developing signed language technologies to ensure they address the community's needs and will be adopted, and that they do not rely on misconceptions or inaccuracies about signed language. Hearing researchers cannot relate to the deaf experience or fully understand the context in which the tools being developed would be used, nor can they speak for the deaf. Therefore, we encourage the creation of a long-term collaborative environment between signed language researchers and users, so that deaf users can identify meaningful challenges, and provide insights on the considerations to take, while researchers cater to the signers' needs as the field evolves. We also recommend reaching out to signing communities for reviewing papers on signed languages, to ensure an adequate evaluation of this type of research results published at ACL venues. There are several ways to connect with Deaf communities for collaboration: one can seek deaf students in their local community, reach out to schools for the deaf, contact deaf linguists, join a network of sign-related researchers[3], and participate in deaf-led projects.

5 Conclusions

We urge the inclusion of signed languages in NLP. We believe that the NLP community is well-positioned, especially with the plethora of successful spoken language processing methods coupled with the recent advent of computer vision tools for videos, to bring the linguistic insight needed for better signed language models. We hope to see an increase in both the interests and efforts in collecting signed language resources and developing signed language tools while building a strong collaboration with signing communities and related fields in computer science.

References

[Borg and Camilleri, 2019] Mark Borg and Kenneth P Camilleri. Sign language detection "in the wild" with recurrent neural networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1637–1641. IEEE, 2019.

[Bull *et al.*, 2020] Hannah Bull, Michèle Gouiffès, and Annelies Braffort. Automatic segmentation of sign language into subtitle-units. In *European Conference on Computer Vision*, pages 186–198. Springer, 2020.

[Camgöz *et al.*, 2018] Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7784–7793, 2018.

[3]https://www.crest-network.com

[Camgöz *et al.*, 2020] Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. Multi-channel transformers for multi-articulatory sign language translation. In *European Conference on Computer Vision*, pages 301–319, 2020.

[Dreuw *et al.*, 2006] Philippe Dreuw, Thomas Deselaers, Daniel Keysers, and Hermann Ney. Modeling image variability in appearance-based gesture recognition. In *ECCV workshop on statistical methods in multi-image and video processing*, pages 7–18, 2006.

[Fenlon *et al.*, 2015] Jordan Fenlon, Kearsy Cormier, and Adam Schembri. Building bsl signbank: The lemma dilemma revisited. *International Journal of Lexicography*, 28(2):169–206, 2015.

[Gebre *et al.*, 2013] Binyam Gebrekidan Gebre, Peter Wittenburg, and Tom Heskes. Automatic sign language identification. In *2013 IEEE International Conference on Image Processing*, pages 2626–2630. IEEE, 2013.

[Hanke *et al.*, 2020] Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. Extending the Public DGS Corpus in size and depth. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages*, 2020.

[Imashev *et al.*, 2020] Alfarabi Imashev, Medet Mukushev, Vadim Kimmelman, and Anara Sandygulova. A dataset for linguistic understanding, visual evaluation, and recognition of sign languages: The k-rsl. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 631–640, 2020.

[Kimmelman, 2014] Vadim Kimmelman. Information structure in russian sign language and sign language of the netherlands. *Sign Language & Linguistics*, 18(1):142–150, 2014.

[Li *et al.*, 2020] Dongxu Li, Cristian Rodriguez, Xin Yu, and Hongdong Li. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 1459–1469, 2020.

[Monteiro *et al.*, 2016] Caio DD Monteiro, Christy Maria Mathew, Ricardo Gutierrez-Osuna, and Frank Shipman. Detecting and identifying sign languages through visual features. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 287–290. IEEE, 2016.

[Moryossef *et al.*, 2021] Amit Moryossef, Kayo Yin, Graham Neubig, and Yoav Goldberg. Data augmentation for sign language gloss translation. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, 2021.

[Ormel and Crasborn, 2012] Ellen Ormel and Onno Crasborn. Prosodic correlates of sentences in signed languages: A literature review and suggestions for new types of studies. *Sign Language Studies*, 12(2):279–315, 2012.

[Padden, 1988] C. Padden. *Interaction of Morphology and Syntax in American Sign Language*. Outstanding Disc Linguistics Series. Garland, 1988.

[Saunders *et al.*, 2020] Ben Saunders, Necati Cihan Camgöz, and Richard Bowden. Everybody sign now: Translating spoken language to photo realistic sign language video. *arXiv preprint arXiv:2011.09846*, 2020.

[Shi *et al.*, 2018] B. Shi, A. Martinez Del Rio, J. Keane, J. Michaux, G. Shakhnarovich D. Brentari, and K. Livescu. American sign language fingerspelling recognition in the wild. *SLT*, December 2018.

[Sincan and Keles, 2020] Ozge Mercanoglu Sincan and Hacer Yalim Keles. Autsl: A large scale multi-modal turkish sign language dataset and baseline methods. *IEEE Access*, 8:181340–181355, 2020.

[Sohn *et al.*, 1999] Jongseo Sohn, Nam Soo Kim, and Wonyong Sung. A statistical model-based voice activity detection. *IEEE signal processing letters*, 6(1):1–3, 1999.

[Stokoe, 1960] Jr. Stokoe, William C. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *The Journal of Deaf Studies and Deaf Education*, 10(1):3–37, 01 1960.

[Stoll *et al.*, 2020] Stephanie Stoll, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. Text2sign: towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, pages 1–18, 2020.

[Yin and Read, 2020a] Kayo Yin and Jesse Read. Attention is all you sign: sign language translation with transformers. In *Sign Language Recognition, Translation and Production (SLRTP) Workshop-Extended Abstracts*, volume 4, 2020.

[Yin and Read, 2020b] Kayo Yin and Jesse Read. Better sign language translation with STMC-transformer. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5975–5989, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.

[Yin *et al.*, 2021a] Kayo Yin, Kenneth DeHaan, and Malihe Alikhani. Signed coreference resolution. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4950–4961, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.

[Yin *et al.*, 2021b] Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. Including signed languages in natural language processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7347–7360, 2021.