

Including Signed Languages in Natural Language Processing

Kayo Yin¹, Amit Moryossef², Julie Hochgesang³, Yoav Goldberg^{2,4}, Malihe Alikhani⁵

¹Language Technologies Institute, Carnegie Mellon University

²Bar-Ilan University

³Department of Linguistics, Gallaudet University

⁴Allen Institute for AI

⁵School of Computing and Information, University of Pittsburgh

kayoy@cs.cmu.edu, amitmoryossef@gmail.com

julie.hochgesang@gallaudet.edu, yogo@cs.biu.ac.il, 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 the direction of 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 people¹ from modern language technologies.

¹According to World Federation of the Deaf
<https://wfdeaf.org/our-work/>

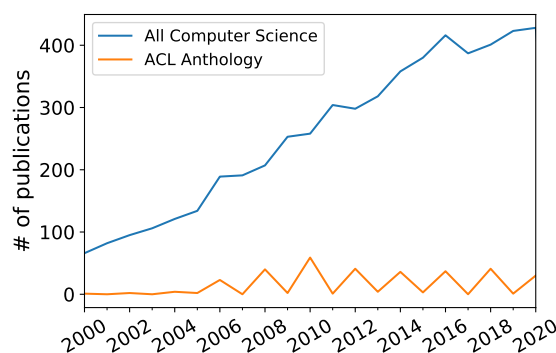


Figure 1: Evolution of the number of publications referring to sign language in its title from computer science venues and in the ACL anthology. Publications in computer science are extracted from the Semantic Scholar archive (Ammar et al., 2018).

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 (Padden and Humphries, 1988; Glickman and Hall, 2018). 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 (Figure 1). 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 (§3), and current SLP techniques fail to address or leverage the linguistic structure of signed languages (§4). 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 (Harris et al., 2009).

After identifying the challenges and open problems to successfully include signed languages in NLP (§5), 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.

2 Background and Related Work

2.1 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 as a research area. Nevertheless, antiquated notions that deprioritized signed languages continue to do harm and subjects many to linguistic neglect (Humphries et al., 2016). 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 (Murray et al., 2020). This language deprivation can lead to life-long consequences on the cognitive, linguistic, socioemotional, and academic development of the deaf (Hall et al., 2017).

Signed languages are the primary languages of communication for the Deaf² and 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.

2.2 Sign Language Processing in the Literature

Jaffe (1994); Ong and Ranganath (2005); Parton

²When capitalized, “Deaf” refers to a community of deaf people who share a language and a culture, whereas the lower-case “deaf” refers to the audiological condition of not hearing.

(2006) survey early works in SLP that were mostly limited to using sensors to capture fingerspelling and isolated signs, or use rules to synthesize signs from spoken language text, due to the lack of adequate CV technology at the time to process videos. This paper will instead focus on more recent vision-based and data-driven approaches that are non-intrusive and more powerful. The introduction of a continuous signed language benchmark dataset (Forster et al., 2014; Cihan Camgöz et al., 2018), coupled with the advent of deep learning for visual processing, lead to increased efforts to recognize signed expressions from videos. Recent surveys on SLP mostly review these different approaches for sign language recognition developed by the CV community (Koller, 2020; Rastgoo et al., 2020; Adaloglou et al., 2020).

Meanwhile, signed languages have remained relatively overlooked in NLP literature (Figure 1). Bragg et al. (2019) argue the importance of an interdisciplinary approach to SLP, raising the importance of NLP involvement among other disciplines. We take this argument further by diving into the linguistic modeling challenges for signed languages and providing a roadmap of open questions to be addressed by the NLP community, in hopes of stimulating efforts from an NLP perspective towards research on signed languages.

3 Sign Language Linguistics

Signed languages consist of phonological, morphological, syntactic, and semantic levels of structure that fulfill the same social, cognitive, and communicative purposes as other natural languages. While spoken languages primarily channel the oral-auditory modality, signed languages use the visual-gestural modality, relying on the face, hands, body of the signer, and the space around them to create distinctions in meaning. We present the linguistic features of signed languages³ that must be taken into account during their modeling.

Phonology Signs are composed of minimal units that combine manual features such as hand configuration, palm orientation, placement, contact, path movement, local movement, as well as non-manual features including eye aperture, head movement, and torso positioning⁴ (Liddell and Johnson, 1989;

³We mainly refer to ASL, where most sign language research has been conducted, but not exclusively.

⁴In this work, we focus on visual signed languages rather than tactile systems such as Pro-Tactile ASL which DeafBlind

Johnson and Liddell, 2011; Brentari, 2011; Sandler, 2012). In both signed and spoken languages, not all possible phonemes are realized, and inventories of two languages' phonemes/features may not overlap completely. Different languages are also subject to rules for the allowed combinations of features.

Simultaneity Though an ASL sign takes about twice as long to produce than an English word, the rates of transmission of information between the two languages are similar (Bellugi and Fischer, 1972). One way signed languages compensate for the slower production rate of signs is through simultaneity: signed languages make use of multiple visual cues to convey different information simultaneously (Sandler, 2012). For example, the signer may produce the sign for 'cup' on one hand while simultaneously pointing to the actual cup with the other to express "that cup". Similarly to tone in spoken languages, the face and torso can convey additional affective information (Liddell et al., 2003; Johnston and Schembri, 2007). Facial expressions can modify adjectives, adverbs, and verbs; a head shake can negate a phrase or sentence; eye direction can help indicate referents.

Referencing The signer can introduce referents in discourse either by pointing to their actual locations in space, or by assigning a region in the signing space to a non-present referent and by pointing to this region to refer to it (Rathmann and Mathur, 2011; Schembri et al., 2018). Signers can also establish relations between referents grounded in signing space by using directional signs or embodying the referents using body shift or eye gaze (Dudis, 2004; Liddell and Metzger, 1998). Spatial referencing also impacts morphology when the directionality of a verb depends on the location of the reference to its subject and/or object (de Beuzeville, 2008; Fenlon et al., 2018): for example, a directional verb can move from the location of its subject and ending at the location of its object. While the relation between referents and verbs in spoken language is more arbitrary, referent relations are usually grounded in signed languages. The visual space is heavily exploited to make referencing clear.

Another way anaphoric entities are referenced in sign language is by using classifiers or depicting signs (Supalla, 1986; Wilcox and Hafer, 2004; Roy, 2011) that help describe the characteristics of the

Americans sometimes prefer.

referent. Classifiers are typically one-handed signs that do not have a particular location or movement assigned to them, or derive features from meaningful discourse (Liddell et al., 2003), so they can be used to convey how the referent relates to other entities, describe its movement, and give more details. For example, to tell about a car swerving and crashing, one might use the hand classifier for a vehicle, move it to indicate swerving, and crash it with another entity in space.

To quote someone other than oneself, signers perform *role shift* (Cormier et al., 2015), where they may physically shift in space to mark the distinction, and take on some characteristics of the people they are representing. For example, to recount a dialogue between a taller and a shorter person, the signer may shift to one side and look up when taking the shorter person’s role, shift to the other side and look down when taking the taller person’s role.

Fingerspelling Fingerspelling is a result of language contact between a signed language and a surrounding spoken language written form (Battison, 1978; Wilcox, 1992; Brentari and Padden, 2001; Patrie and Johnson, 2011). A set of manual gestures correspond with a written orthography or phonetic system. Fingerspelling is often used to indicate names or places or new concepts from the spoken language but often have become integrated into the signed languages themselves as another linguistic strategy (Padden, 1998; Montemurro and Brentari, 2018).

4 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.

4.1 Representations of Signed Languages

Representation is a significant challenge for SLP, as unlike spoken languages, signed languages have no widely adopted written form. Figure 2 illustrates each signed language representation we will describe below.

Videos are the most straightforward representation of a signed language and can amply incorporate the information conveyed through sign. One major drawback of using videos is their high dimensionality: they usually include more information than needed for modeling, and are expensive to store, transmit, and encode. As facial features are

essential in sign, anonymizing raw videos also remains an open problem, limiting the possibility of making these videos publicly available (Isard, 2020).

Poses reduce the visual cues from videos to skeleton-like wireframe or mesh representing the location of joints. While motion capture equipment can often provide better quality pose estimation, it is expensive and intrusive, and estimating pose from videos is the preferred method currently (Pishchulin et al., 2012; Chen et al., 2017; Cao et al., 2019; Güler et al., 2018). Compared to video representations, *accurate* poses are lower in complexity and anonymized, while observing relatively low information loss. However, they remain a continuous, multidimensional representation that is not adapted to most NLP models.

Written notation systems represent signs as discrete visual features. Some systems are written linearly and others use graphemes in two dimensions. While various universal (Sutton, 1990; Prillwitz and Zienert, 1990) and language-specific notation systems (Stokoe Jr, 2005; Kakumasu, 1968; Bergman, 1979) have been proposed, no writing system has been adopted widely by any sign language community, and the lack of standard hinders the exchange and unification of resources and applications between projects. Figure 2 depicts two universal notation systems: SignWriting (Sutton, 1990), a two-dimensional pictographic system, and HamNoSys (Prillwitz and Zienert, 1990), a linear stream of graphemes that was designed to be readable by machines.

Glossing is the transcription of signed languages sign-by-sign, where every sign has a unique identifier. While various sign language corpus projects have provided gloss annotation guidelines (Mesch and Wallin, 2015; Johnston and De Beuzeville, 2016; Konrad et al., 2018), again, there is no single agreed-upon standard. Linear gloss annotations are also an imprecise representation of signed language: they do not adequately capture all information expressed simultaneously through different cues (i.e. body posture, eye gaze) or spatial relations, which leads to an inevitable information loss up to a semantic level that affects downstream performance on SLP tasks (Yin and Read, 2020b).

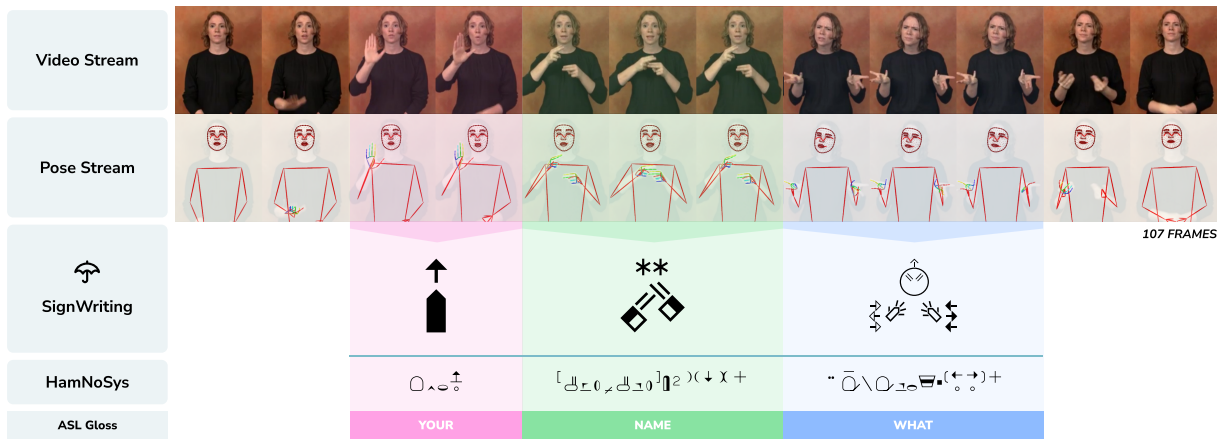


Figure 2: Representations of an American Sign Language phrase with video frames, pose estimations, SignWriting, HamNoSys and glosses. English translation: “What is your name?”

4.2 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 (Mesch and Wallin, 2012; Fenlon et al., 2015; Crasborn et al., 2016; Gutierrez-Sigut et al., 2016) map a spoken language word or short phrase to a signed language video. One notable dictionary is, SpreadTheSign⁵ 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, 2019). However, they only capture one aspect of signed languages.

Isolated sign corpora are collections of annotated single signs. They are synthesized (Ebling et al., 2018; Huang et al., 2018; Sincan and Keles, 2020; Hassan et al., 2020) or mined from online resources (Vaezi Joze and Koller, 2019; 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 (Arivazhagan et al., 2019). Moreover, while automatic speech recognition (ASR) datasets contain up to 50,000 hours of recordings (Pratap et al., 2020), 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 (Databases, 2007; Crasborn and Zwitterlood, 2008; Ko et al., 2019; Hanke et al., 2020) or recorded in studio conditions (Forster et al., 2014; Cihan Camgöz et al., 2018), which does not account for noise in real-life conditions. Moreover, some contain signed interpretations of spoken language rather than naturally-produced signs, which may not accurately represent native signing since translation is now a part of the discourse event.

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

⁵<https://www.spreadthesign.com/>

in signing videos, limiting its open distribution, and developing anonymization with minimal information loss, or accurate anonymous representations is a promising research problem.

4.3 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; Moryossef et al., 2020) 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 detection (VAD) (Sohn et al., 1999; Ramirez et al., 2004), 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 (McKee and Kennedy, 2000; Kimmelman, 2014; Ferreira-Brito, 1984; Shroyer and Shroyer, 1984) which have not been explored for this task.

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 (Santemiz et al., 2009; Farag and Brock, 2019; 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 expres-

sions, eye apertures) (Sandler, 2010; Ormel and Crasborn, 2012).

Recognition Sign language recognition (SLR) detects and label signs from a video, either on isolated (Imashev et al., 2020; Sincan and Keles, 2020) or continuous (Cui et al., 2017; Camgöz et al., 2018, 2020b) 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, 2020b; Yin and Read, 2020a,b; Moryossef et al., 2021) or on pose estimations and sign articulators from videos (Ko et al., 2019; Camgöz et al., 2020a), 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., 2018, 2020), while more recent methods (Saunders et al., 2020b,a; Zelinka and Kanis, 2020; Xiao 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.

5 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, the development of known tasks in the standard NLP pipeline to 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 signed language technologies can benefit signers, all the while respecting the Deaf community's ownership of signed languages.

5.1 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, all the while considering: how do we define lexical units in signed languages? (Johnston and Schembri, 1999; Johnston, 2010) To what degree can phonological units of signed languages be mapped to lexical units? Should we model the articulators of signs separately or together? What are the cross-linguistic phonological differences to consider? To what extent can ideas used in automatic speech recognition be applied to signed languages?

Syntactic Analysis Part-of-speech (POS) tagging and syntactic parsing are fundamental to understand the meaning of words in context. Yet, no such linguistic tools for automatic syntactic analyses exist. To develop such tools, we must first define to what extent POS tagging and syntactic parsing for spoken languages also generalize to signed languages - do we need a new set of POS and dependency tags for signed languages? How are morphological features expressed? What are the annotation guidelines to create datasets on syntax? Can we draw on linguistic theories to design features and rules that perform these tasks? Are there typologically similar spoken languages for some signed languages we can perform transfer learning with?

Named Entity Recognition (NER) Recognizing named entities and finding relationships between them are highly important in information retrieval and classification. Named entities in signed languages can be produced by a finger-spelled sequence, a sign, or even through mouthing of the name while the referent is introduced through pointing. Bleicken et al. (2016) attempt NER in German Sign Language (DGS) to perform anonymization, but only do so indirectly, by either performing NER on the gold DGS gloss annotations and German translations or manually on the videos. We instead propose NER in a fully automated fashion while considering, what are the visual markers of named entities? How are they introduced and referenced? How are relationships between them established?

Coreference Resolution Resolving coreference is crucial for language understanding. In signed languages, present referents, where the signer explicitly points to the entity in question, are relatively unambiguous. In contrast, non-present referents and classifiers are heavily grounded in the signing space, so good modeling of the spatial coherence in sign language is required. Evidence suggests that classic theoretical frameworks, such as discourse representation theory, may extend to signed languages (Steinbach and Onea, 2016). We pose the following questions: to what extent can automatic coreference resolution of spoken languages be applied to signed languages? How do we keep track of referents in space? How can we leverage spatial relations to resolve ambiguity?

Towards Linguistically Informed and Multimodal SLP We highly encourage the collaboration of multimodal and SLP research communities to develop powerful SLP models informed by core NLP tools such as the ones discussed, all the while processing and relating information from both linguistic and visual modalities. On the one hand, theories and methods to reason multimodal messages can enhance the joint modeling of vision and language in signed languages. SLP is especially subject to three of the core technical challenges in multimodal machine learning (Baltrušaitis et al., 2018): **translation** - how do we map visual-gestural information to/from audio-oral and textual information? **alignment** - how do we relate signed language units to spoken language units? **co-learning** - can we transfer high-resource spoken language knowledge to signed language? On the other hand, meaning in spoken languages is not only conveyed through speech or text but also through the visual modality. Studying signed languages can give a better understanding of how to model co-speech gestures, spatial discourse relations, and conceptual grounding of language through vision.

5.2 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 ac-

curately. What constitutes an ideal signed language dataset is an open question, we suggest including 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.

To illustrate the importance of data quality during modeling, we first take as an example a current benchmark for SLP, the RWTH-PHOENIX-Weather 2014T dataset (Cihan Camgöz et al., 2018) of German Sign Language, that does not meet most of the above criteria: it is restricted to the weather domain (1); contains only around 8K segments with 1K unique signs (2); filmed in studio conditions (3); interpreted from German utterances (4); and signed by nine Caucasian interpreters (5,6). Although this dataset successfully addressed data scarcity issues at the time and successfully rendered results comparable and fueled competitive research, it does not accurately represent signed languages in the real world. On the other hand, the Public DGS Corpus (Hanke et al., 2020) is an open-domain (1) dataset consisting of 50 hours of natural signing (4) by 330 native signers from various regions in Germany (5,6), annotated with glosses, HamNoSys and German translations (7), meeting all but two requirements we suggest.

We train a gloss-to-text sign language translation transformer (Yin and Read, 2020b) on both datasets. On RWTH-PHOENIX-Weather 2014T, we obtain **22.17** BLEU on testing; on Public DGS Corpus, we obtain a mere **3.2** BLEU. Although Transformers achieve encouraging results on RWTH-PHOENIX-Weather 2014T (Saunders et al., 2020b; Camgöz et al., 2020a), they fail on more realistic, open-domain data. These results reveal that firstly, for real-world applications, we need more data to train such types of models, and secondly, while available data is severely limited in size, less data-hungry and more linguistically-informed approaches may be more suitable. This experiment reveals how it is crucial to use data that accurately represent the complexity and diversity of signed languages to precisely assess what types of methods are suitable, and how well our models would deploy to the real world.

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 require 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.

5.3 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 (Bird, 2020).

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 (Erard, 2017). 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 (Harris et al., 2009; Kusters et al., 2017). 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 researchers of sign-related technologies⁶, and/or participate in deaf-led projects.

6 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.

Acknowledgements

We would like to thank Marc Schulder, Claude Mauk, David Mortensen, Chaitanya Ahuja, Sidharth Dalmia, Shruti Palaskar and Graham Neubig as well as the anonymous reviewers for their helpful feedback and insightful discussions.

⁶<https://www.crest-network.com/>

References

- Nikolas Adaloglou, Theocharis Chatzis, Ilias Papatratis, Andreas Stergioulas, Georgios Th Papadopoulos, Vassia Zacharopoulou, George J Xydopoulos, Klimnis Atzakas, Dimitris Papazachariou, and Petros Daras. 2020. A comprehensive study on sign language recognition methods. *arXiv preprint arXiv:2007.12530*.
- Waleed Ammar, Dirk Groeneveld, Chandra Bhagavathula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu Ha, Rodney Kinney, Sebastian Kohlmeier, Kyle Lo, Tyler Murray, Hsu-Han Ooi, Matthew Peters, Joanna Power, Sam Skjonsberg, Lucy Wang, Chris Wilhelm, Zheng Yuan, Madeleine van Zuylen, and Oren Etzioni. 2018. **Construction of the literature graph in semantic scholar**. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*, pages 84–91, New Orleans - Louisiana. Association for Computational Linguistics.
- Naveen Arivazhagan, Ankur Bapna, Orhan Firat, Dmitry Lepikhin, Melvin Johnson, Maxim Krikun, Mia Xu Chen, Yuan Cao, George Foster, Colin Cherry, et al. 2019. Massively multilingual neural machine translation in the wild: Findings and challenges. *arXiv preprint arXiv:1907.05019*.
- Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2018. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):423–443.
- Robbin Battison. 1978. *Lexical borrowing in American sign language*. ERIC.
- Ursula Bellugi and Susan Fischer. 1972. A comparison of sign language and spoken language. *Cognition*, 1(2-3):173–200.
- Brita Bergman. 1979. *Signed Swedish*. National Swedish Board of Education [Skolöverstyr:].
- Louise de Beuzeville. 2008. Pointing and verb modification: the expression of semantic roles in the auslan corpus. In *Workshop Programme*, page 13. Citeseer.
- Steven Bird. 2020. **Decolonising speech and language technology**. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Julian Bleicken, Thomas Hanke, Uta Salden, and Sven Wagner. 2016. **Using a language technology infrastructure for German in order to anonymize German Sign Language corpus data**. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3303–3306, Portorož, Slovenia. European Language Resources Association (ELRA).
- Mark Borg and Kenneth P Camilleri. 2019. 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.
- Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. **Sign language recognition, generation, and translation: An interdisciplinary perspective**. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '19*, page 16–31, New York, NY, USA. Association for Computing Machinery.
- Diane Brentari. 2011. Sign language phonology. *The handbook of phonological theory*, pages 691–721.
- Diane Brentari and Carol Padden. 2001. A language with multiple origins: Native and foreign vocabulary in american sign language. *Foreign vocabulary in sign language: A cross-linguistic investigation of word formation*, pages 87–119.
- Hannah Bull, Michèle Gouiffès, and Annelies Braffort. 2020. Automatic segmentation of sign language into subtitle-units. In *European Conference on Computer Vision*, pages 186–198. Springer.
- Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7784–7793.
- Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020a. Multi-channel transformers for multi-articulatory sign language translation. In *European Conference on Computer Vision*, pages 301–319.
- Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020b. Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10023–10033.
- Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Yu Chen, Chunhua Shen, Xiu-Shen Wei, Lingqiao Liu, and Jian Yang. 2017. Adversarial posenet: A structure-aware convolutional network for human pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1212–1221.
- Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the*

- IEEE Conference on Computer Vision and Pattern Recognition*, pages 7784–7793.
- Kearsy Cormier, Sandra Smith, and Zed Sevcikova-Sehry. 2015. Rethinking constructed action. *Sign Language & Linguistics*, 18(2):167–204.
- Onno Crasborn, R Bank, I Zwitserlood, E Van der Kooij, E Ormel, J Ros, A Schüller, A de Meijer, M van Zuilen, YE Nauta, et al. 2016. Ngt signbank. *Nijmegen: Radboud University, Centre for Language Studies*.
- Onno A Crasborn and IEP Zwitserlood. 2008. The corpus ngt: an online corpus for professionals and laymen.
- Runpeng Cui, Hu Liu, and Changshui Zhang. 2017. Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7361–7369.
- NCSLGR Databases. 2007. Volumes 2–7.
- Philippe Dreuw, Thomas Deselaers, Daniel Keysers, and Hermann Ney. 2006. Modeling image variability in appearance-based gesture recognition. In *ECCV workshop on statistical methods in multi-image and video processing*, pages 7–18.
- Paul G Dudis. 2004. Body partitioning and real-space blends. *Cognitive linguistics*, 15(2):223–238.
- Sarah Ebling, Necati Cihan Camg ö z, Penny Boyes Braem, Katja Tissi, Sandra Sidler-Miserez, Stephanie Stoll, Simon Hadfield, Tobias Haug, Richard Bowden, Sandrine Tornay, Marzieh Razavi, and Mathew Magimai-Doss. 2018. **SMILE Swiss German sign language dataset**. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Michael Erard. 2017. Why sign-language gloves don’t help deaf people. *The Atlantic*, 9.
- Iva Farag and Heike Brock. 2019. Learning motion disfluencies for automatic sign language segmentation. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7360–7364. IEEE.
- Jordan Fenlon, Kearsy Cormier, and Adam Schembri. 2015. Building bsl signbank: The lemma dilemma revisited. *International Journal of Lexicography*, 28(2):169–206.
- Jordan Fenlon, Adam Schembri, and Kearsy Cormier. 2018. Modification of indicating verbs in british sign language: A corpus-based study. *Language*, 94(1):84–118.
- Lucinda Ferreira-Brito. 1984. Similarities & differences in two brazilian sign languages. *Sign language studies*, 42:45–56.
- Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney. 2014. Extensions of the sign language recognition and translation corpus rwth-phoenix-weather. In *LREC*, pages 1911–1916.
- Binyam Gebrekidan Gebre, Peter Wittenburg, and Tom Heskes. 2013. Automatic sign language identification. In *2013 IEEE International Conference on Image Processing*, pages 2626–2630. IEEE.
- Neil S Glickman and Wyatt C Hall. 2018. *Language deprivation and deaf mental health*. Routledge.
- Rıza Alp Güler, Natalia Neverova, and Iasonas Kokkinos. 2018. Densepose: Dense human pose estimation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7297–7306.
- Eva Gutierrez-Sigut, Brendan Costello, Cristina Baus, and Manuel Carreiras. 2016. Lse-sign: A lexical database for spanish sign language. *Behavior Research Methods*, 48(1):123–137.
- Wyatte C Hall, Leonard L Levin, and Melissa L Anderson. 2017. Language deprivation syndrome: A possible neurodevelopmental disorder with sociocultural origins. *Social psychiatry and psychiatric epidemiology*, 52(6):761–776.
- Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020. **Extending the Public DGS Corpus in size and depth**. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 75–82, Marseille, France. European Language Resources Association (ELRA).
- Raychelle Harris, Heidi M Holmes, and Donna M Mertens. 2009. Research ethics in sign language communities. *Sign Language Studies*, 9(2):104–131.
- Saad Hassan, Larwan Berke, Elahe Vahdani, Longlong Jing, Yingli Tian, and Matt Huenerfauth. 2020. **An isolated-signing RGBD dataset of 100 American Sign Language signs produced by fluent ASL signers**. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 89–94, Marseille, France. European Language Resources Association (ELRA).
- Jie Huang, Wengang Zhou, Qilin Zhang, Houqiang Li, and Weiping Li. 2018. Video-based sign language recognition without temporal segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

- Tom Humphries, Poorna Kushalnagar, Gaurav Mathur, Donna Jo Napoli, Carol Padden, Christian Rathmann, and Scott Smith. 2016. Avoiding linguistic neglect of deaf children. *Social Service Review*, 90(4):589–619.
- Alfarabi Imashev, Medet Mukushev, Vadim Kimmelman, and Anara Sandygulova. 2020. 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.
- Amy Isard. 2020. Approaches to the anonymisation of sign language corpora. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 95–100.
- David L Jaffe. 1994. Evolution of mechanical fingerspelling hands for people who are deaf-blind. *Journal of rehabilitation research and development*, 31(3):236–244.
- Robert E Johnson and Scott K Liddell. 2011. Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies*, 11(2):241–274.
- Trevor Johnston. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International journal of corpus linguistics*, 15(1):106–131.
- Trevor Johnston and Louise De Beuzeville. 2016. Auslan corpus annotation guidelines. *Auslan Corpus*.
- Trevor Johnston and Adam Schembri. 2007. *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge University Press.
- Trevor Johnston and Adam C Schembri. 1999. On defining lexeme in a signed language. *Sign language & linguistics*, 2(2):115–185.
- Jim Kakumasu. 1968. Urubu sign language. *International journal of American linguistics*, 34(4):275–281.
- Vadim Kimmelman. 2014. Information structure in russian sign language and sign language of the netherlands.
- Sang-Ki Ko, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. 2019. Neural sign language translation based on human keypoint estimation. *Applied Sciences*, 9(13):2683.
- Oscar Koller. 2020. Quantitative survey of the state of the art in sign language recognition. *arXiv preprint arXiv:2008.09918*.
- Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2018. Public dgs corpus: Annotation conventions. Technical report, Project Note AP03-2018-01, DGS-Korpus project, IDGS, Hamburg University.
- Annelies Kusters, Maartje De Meulder, and Dai O’Brien. 2017. *Innovations in deaf studies: The role of deaf scholars*. Oxford University Press.
- Dongxu Li, Cristian Rodriguez, Xin Yu, and Hongdong Li. 2020. 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.
- Scott K Liddell and Robert E Johnson. 1989. American sign language: The phonological base. *Sign language studies*, 64(1):195–277.
- Scott K Liddell and Melanie Metzger. 1998. Gesture in sign language discourse. *Journal of pragmatics*, 30(6):657–697.
- Scott K Liddell et al. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge University Press.
- David McKee and Graeme Kennedy. 2000. Lexical comparison of signs from american, australian, british and new zealand sign languages. *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, pages 49–76.
- Johanna Mesch and Lars Wallin. 2012. From meaning to signs and back: Lexicography and the swedish sign language corpus. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 123–126.
- Johanna Mesch and Lars Wallin. 2015. Gloss annotations in the swedish sign language corpus. *International Journal of Corpus Linguistics*, 20(1):102–120.
- Caio DD Monteiro, Christy Maria Mathew, Ricardo Gutierrez-Osuna, and Frank Shipman. 2016. Detecting and identifying sign languages through visual features. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 287–290. IEEE.
- Kathryn Montemurro and Diane Brentari. 2018. Emphatic fingerspelling as code-mixing in american sign language. *Proceedings of the Linguistic Society of America*, 3(1):61–1.
- Amit Moryossef, Ioannis Tsochantaridis, Roei Aharoni, Sarah Ebling, and Srini Narayanan. 2020. Real-time sign language detection using human pose estimation. In *European Conference on Computer Vision*, pages 237–248. Springer.

- Amit Moryossef, Kayo Yin, Graham Neubig, and Yoav Goldberg. 2021. Data augmentation for sign language gloss translation. *arXiv preprint arXiv:2105.07476*.
- Joseph J Murray, Wyatt C Hall, and Kristin Snoddon. 2020. The importance of signed languages for deaf children and their families. *The Hearing Journal*, 73(3):30–32.
- Sylvie CW Ong and Surendra Ranganath. 2005. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Computer Architecture Letters*, 27(06):873–891.
- Ellen Ormel and Onno Crasborn. 2012. Prosodic correlates of sentences in signed languages: A literature review and suggestions for new types of studies. *Sign Language Studies*, 12(2):279–315.
- C. Padden. 1988. *Interaction of Morphology and Syntax in American Sign Language*. Outstanding Disc Linguistics Series. Garland.
- Carol A Padden. 1998. The asl lexicon. *Sign language & linguistics*, 1(1):39–60.
- Carol A Padden and Tom Humphries. 1988. *Deaf in America*. Harvard University Press.
- Becky Sue Parton. 2006. Sign language recognition and translation: A multidisciplinary approach from the field of artificial intelligence. *Journal of deaf studies and deaf education*, 11(1):94–101.
- Carol J Patrie and Robert E Johnson. 2011. *Finger-spelled word recognition through rapid serial visual presentation: RSVP*. DawnSignPress.
- Leonid Pishchulin, Arjun Jain, Mykhaylo Andriluka, Thorsten Horn, and Bernt Schiele. 2012. Articulated people detection and pose estimation: Reshaping the future. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3178–3185. IEEE.
- Vineel Pratap, Qiantong Xu, Anuroop Sriram, Gabriel Synnaeve, and Ronan Collobert. 2020. **MLS: A Large-Scale Multilingual Dataset for Speech Research**. In *Proc. Interspeech 2020*, pages 2757–2761.
- Siegmund Prillwitz and Heiko Zienert. 1990. Hamburg notation system for sign language: Development of a sign writing with computer application. In *Current trends in European Sign Language Research. Proceedings of the 3rd European Congress on Sign Language Research*, pages 355–379.
- Javier Ramirez, José C Segura, Carmen Benitez, Angel De La Torre, and Antonio Rubio. 2004. Efficient voice activity detection algorithms using long-term speech information. *Speech communication*, 42(3-4):271–287.
- Razieh Rastgoo, Kourosh Kiani, and Sergio Escalera. 2020. Sign language recognition: A deep survey. *Expert Systems with Applications*, page 113794.
- Christian Rathmann and Gaurav Mathur. 2011. A featural approach to verb agreement in signed languages. *Theoretical Linguistics*, 37(3-4):197–208.
- Cynthia B Roy. 2011. *Discourse in signed languages*. Gallaudet University Press.
- Wendy Sandler. 2010. Prosody and syntax in sign languages. *Transactions of the philological society*, 108(3):298–328.
- Wendy Sandler. 2012. The phonological organization of sign languages. *Language and linguistics compass*, 6(3):162–182.
- Pinar Santemiz, Oya Aran, Murat Saraclar, and Lale Akarun. 2009. Automatic sign segmentation from continuous signing via multiple sequence alignment. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 2001–2008. IEEE.
- Ben Saunders, Necati Cihan Camgöz, and Richard Bowden. 2020a. Everybody sign now: Translating spoken language to photo realistic sign language video. *arXiv preprint arXiv:2011.09846*.
- Ben Saunders, Necati Cihan Camgöz, and Richard Bowden. 2020b. Progressive transformers for end-to-end sign language production. In *European Conference on Computer Vision*, pages 687–705.
- Adam Schembri, Kearsy Cormier, and Jordan Fenlon. 2018. Indicating verbs as typologically unique constructions: Reconsidering verb ‘agreement’ in sign languages. *Glossa: a journal of general linguistics*, 3(1).
- B. Shi, A. Martinez Del Rio, J. Keane, D. Brentari, G. Shakhnarovich, and K. Livescu. 2019. Fingerspelling recognition in the wild with iterative visual attention. *ICCV*.
- B. Shi, A. Martinez Del Rio, J. Keane, J. Michaux, G. Shakhnarovich D. Brentari, and K. Livescu. 2018. American sign language fingerspelling recognition in the wild. *SLT*.
- Edgar H Shroyer and Susan P Shroyer. 1984. *Signs across America: A look at regional differences in American Sign Language*. Gallaudet University Press.
- Ozge Mercanoglu Sincan and Hacer Yalim Keles. 2020. Autsl: A large scale multi-modal turkish sign language dataset and baseline methods. *IEEE Access*, 8:181340–181355.
- Jongseo Sohn, Nam Soo Kim, and Wonyong Sung. 1999. A statistical model-based voice activity detection. *IEEE signal processing letters*, 6(1):1–3.

- Markus Steinbach and Edgar Onea. 2016. A DRT analysis of discourse referents and anaphora resolution in sign language. *Journal of Semantics*, 33(3):409–448.
- Jr. Stokoe, William C. 1960. [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.
- William C Stokoe Jr. 2005. Sign language structure: An outline of the visual communication systems of the american deaf. *Journal of deaf studies and deaf education*, 10(1):3–37.
- Stephanie Stoll, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. 2018. Sign language production using neural machine translation and generative adversarial networks. In *Proceedings of the 29th British Machine Vision Conference (BMVC 2018)*. British Machine Vision Association.
- Stephanie Stoll, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. 2020. Text2sign: towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*, pages 1–18.
- Ted Supalla. 1986. The classifier system in american sign language. *Noun classes and categorization*, 7:181–214.
- Valerie Sutton. 1990. *Lessons in sign writing*. Sign-Writing.
- Hamid Vaezi Joze and Oscar Koller. 2019. [Ms-asl: A large-scale data set and benchmark for understanding american sign language](#). In *The British Machine Vision Conference (BMVC)*.
- Sherman Wilcox. 1992. *The phonetics of finger-spelling*, volume 4. John Benjamins Publishing.
- Sherman Wilcox and Sarah Hafer. 2004. Rethinking classifiers. emmorey, k.(ed.).(2003). perspectives on classifier constructions in sign languages. mahwah, nj: Lawrence erlbaum associates. 332 pages. hardcover.
- Qinkun Xiao, Minying Qin, and Yuting Yin. 2020. Skeleton-based chinese sign language recognition and generation for bidirectional communication between deaf and hearing people. *Neural Networks*, 125:41–55.
- Kayo Yin and Jesse Read. 2020a. Attention is all you sign: sign language translation with transformers. In *Sign Language Recognition, Translation and Production (SLRTP) Workshop-Extended Abstracts*, volume 4.
- Kayo Yin and Jesse Read. 2020b. [Better sign language translation with STMC-transformer](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5975–5989, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Jan Zelinka and Jakub Kanis. 2020. Neural sign language synthesis: Words are our glosses. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3395–3403.