Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective

Danielle Bragg[1] **Oscar Koller**[2] **Mary Bellard**[2] **Larwan Berke**[3]
Patrick Boudreault[4] **Annelies Braffort**[5] **Naomi Caselli**[6] **Matt Huenerfauth**[3]
Hernisa Kacorri[7] **Tessa Verhoef**[8] **Christian Vogler**[4] **Meredith Ringel Morris**[1]

[1]Microsoft Research - Cambridge, MA USA & Redmond, WA USA {danielle.bragg,merrie}@microsoft.com
[2]Microsoft - Munich, Germany & Redmond, WA USA {oscar.koller,marybel}@microsoft.com
[3]Rochester Institute of Technology - Rochester, NY USA {larwan.berke,matt.huenerfauth}@rit.edu
[4]Gallaudet University - Washington, DC USA {patrick.boudreault,christian.vogler}@gallaudet.edu
[5]LIMSI, CNRS, Paris-Saclay University - Orsay, France annelies.braffort@limsi.fr
[6]Boston University - Boston, MA USA nkc@bu.edu
[7]University of Maryland - College Park, MD USA hernisa@umd.edu
[8]Leiden University - Leiden, Netherlands t.verhoef@liacs.leidenuniv.nl

ABSTRACT

Developing successful sign language recognition, generation, and translation systems requires expertise in a wide range of fields, including computer vision, computer graphics, natural language processing, human-computer interaction, linguistics, and Deaf culture. Despite the need for deep interdisciplinary knowledge, existing research occurs in separate disciplinary silos, and tackles separate portions of the sign language processing pipeline. This leads to three key questions: 1) What does an interdisciplinary view of the current landscape reveal? 2) What are the biggest challenges facing the field? and 3) What are the calls to action for people working in the field? To help answer these questions, we brought together a diverse group of experts for a two-day workshop. This paper presents the results of that interdisciplinary workshop, providing key background that is often overlooked by computer scientists, a review of the state-of-the-art, a set of pressing challenges, and a call to action for the research community.

CCS Concepts

•General and reference → Reference works; •Human-centered computing → Natural language interfaces;

Author Keywords

sign language; recognition, translation, and generation; ASL

INTRODUCTION

Sign language recognition, generation, and translation is a research area with high potential impact. (For brevity, we refer to these three related topics as "sign language processing" throughout this paper.) According to the World Federation of the Deaf, there are over 300 sign languages used around the world, and 70 million deaf people using them [89]. Sign languages, like all languages, are naturally evolved, highly structured systems governed by a set of linguistic rules. They are distinct from spoken languages – i.e., American Sign Language (ASL) is not a manual form of English – and do not have standard written forms. However, the vast majority of communications technologies are designed to support spoken or written language (which excludes sign languages), and most hearing people do not know a sign language. As a result, many communication barriers exist for deaf sign language users.

Sign language processing would help break down these barriers for sign language users. These technologies would make voice-activated services newly accessible to deaf sign language users – for example, enabling the use of personal assistants (e.g., Siri and Alexa) by training them to respond to people signing. They would also enable the use of text-based systems – for example by translating signed content into written queries for a search engine, or automatically replacing displayed text with sign language videos. Other possibilities include automatic transcription of signed content, which would enable indexing and search of sign language videos, real-time interpreting when human interpreters are not available, and many educational tools and applications.

Current research in sign language processing occurs in disciplinary silos, and as a result does not address the problem comprehensively. For example, there are many computer science publications presenting algorithms for recognizing (and less frequently translating) signed content. The teams creating these algorithms often lack Deaf members with lived experience of the problems the technology could or should solve, and lack knowledge of the linguistic complexities of the language for which their algorithms must account. The algorithms are also often trained on datasets that do not reflect real-world use cases. As a result, such single-disciplinary approaches to sign language processing have limited real-world value [39].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ASSETS'19, October 28–30, 2019, Pittsburgh, PA, USA

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ISBN 978-1-4503-6676-2/19/10…$15.00
https://doi.org/10.1145/3308561.3353774

To overcome these problems, we argue for an interdisciplinary approach to sign language processing. Deaf studies must be included in order to understand the community that the technology is built to serve. Linguistics is essential for identifying the structures of sign languages that algorithms must handle. Natural Language Processing (NLP) and machine translation (MT) provide powerful methods for modeling, analyzing, and translating. Computer vision is required for detecting signed content, and computer graphics are required for generating signed content. Finally, Human-Computer Interaction (HCI) and design are essential for creating end-to-end systems that meet the community's needs and integrate into people's lives.

This work addresses the following questions:

Q1: What is the current state of sign language processing, from an interdisciplinary perspective?

Q2: What are the biggest challenges facing the field, from an interdisciplinary perspective?

Q3: What calls to action are there for the field, that resonate across disciplines?

To address these questions, we conducted an interdisciplinary workshop with 39 participants. The workshop brought together academics from diverse backgrounds to synthesize the state-of-the-art in disparate domains, discuss the biggest challenges facing sign language processing efforts, and formulate a call-to-action for the research community. This paper synthesizes the workshop findings, providing a comprehensive interdisciplinary foundation for future research in sign language processing. The audience for this paper includes both newcomers to sign language processing and experts on a portion of the technology seeking to expand their perspective.

The main contributions of this work are:

- orientation and insights for researchers in any domain, in particular those entering the field
- highlighting of needs and opportunities for interdisciplinary collaboration
- prioritization of important problems in the field for researchers to tackle next

BACKGROUND AND RELATED WORK

Building successful sign language processing systems requires an understanding of Deaf culture in order to create systems that align with user needs and desires, and of sign languages to build systems that account for their complex linguistic aspects. Here, we summarize this background, and we also discuss existing reviews of sign language processing, which do not take a comprehensive view of the problem.

Deaf Culture

Sign language users make up cultural minorities, united by common languages and life experience. Many people view deafness not as a disability, but as a cultural identity [47] with many advantages [9]. When capitalized, "Deaf" refers to this cultural identity, while lowercase "deaf" refers to audiological status. Like other cultures, Deaf cultures are characterized by unique sets of norms for interacting and living. Sign languages are a central component of Deaf cultures, their role in Deaf communities even characterized as sacred [6]. Consequently, development of sign language processing systems is highly sensitive, and must do the language justice to gain adoption.

Suppression of sign language communication has been a major form of oppression against the Deaf community. Such discrimination is an example of "audism" [8, 36, 60]. In 1880, an international congress of largely hearing educators of deaf students declared that spoken language should be used for educating deaf children, not sign language [78]. Subsequently, oralism was widely enforced, resulting in training students to lip-read and speak, with varying success. Since then, Deaf communities have fought to use sign languages in schools, work, and public life (e.g., [45]). Linguistic work has helped gain respect for sign languages, by establishing them as natural languages [107]. Legislation has also helped establish legal support for sign language education and use (e.g., [5]). This historical struggle can make development of sign language software particularly sensitive in the Deaf community.

Sign Language Linguistics

Just like spoken languages, sign languages are composed of building blocks, or phonological features, put together under certain rules. The seminal linguistic analysis of a sign language (ASL) revealed that each sign has three main phonological features: handshape, location on the body, and movement [106]. More recent analyses of sign languages offer more sophisticated and detailed phonological analyses [17, 114, 96, 20]. While phonological features are not always meaningful (e.g., the bent index finger in the sign APPLE does not mean anything on its own), they *can* be [18]. For example, in some cases the movement of the sign has a grammatical function. In particular, the direction of movement in verbs can indicate the subject and object of the sentence.

Classifiers represent classes of nouns and verbs – e.g., one handshape in ASL is used for vehicles, another for flat objects, and others for grabbing objects of particular shapes. The vehicle handshape could be combined with a swerving upward movement to mean a vehicle swerving uphill, or a jittery straight movement for driving over gravel. Replacing the handshape could indicate a person walking instead. These handshapes, movements, and locations are not reserved exclusively for classifiers, and can appear in other signs. Recognition software must differentiate between such usages.

Fingerspelling, where a spoken/written word is spelled out using handshapes representing letters, is prevalent in many sign languages. For example, fingerspelling is often used for the names of people or organizations taken from spoken language. Its execution is subject to a high degree of coarticulation, where handshapes change depending on the neighboring letters [67]. Recognition software must be able to identify when a handshape is used for fingerspelling vs. other functionalities.

Sign languages are not entirely expressed with the hands; movement of the eyebrows, mouth, head, shoulders, and eye gaze can all be critical [119, 19]. For example, in ASL, raised eyebrows indicate an open-ended question, and furrowed eye-

brows indicate a yes/no question. Signs can also be modified by adding mouth movements – e.g., executing the sign CUP with different mouth positions can indicate cup size. Sign languages also make extensive use of depiction: using the body to depict an action (e.g., showing how one would fillet a fish), dialogue, or psychological events [34]. Subtle shifts in body positioning and eye gaze can be used to indicate a referent. Sign language recognition software must accurately detect these non-manual components.

There is great diversity in sign language execution, based on ethnicity, geographic region, age, gender, education, language proficiency, hearing status, etc. As in spoken language, different social and geographic communities use different varieties of sign languages (e.g., Black ASL is a distinct dialect of ASL [85]). Unlike spoken languages, sign languages contain enormous variance in fluency. Most deaf children are born to hearing parents, who may not know sign language when the child is born. Consequently, most deaf sign language users learn the language in late childhood or adulthood, typically resulting in lower fluency [84]. Sign language processing software must accurately model and detect this variety, which increases the amount and variety of training data required.

It is difficult to estimate sign language vocabulary size. Existing ASL-to-English dictionaries contain 5-10k signs (e.g., [103]). However, they are not representative of true size, as they lack classifiers, depictions, and other ways signs are modulated to add adjectives, adverbs, and nuanced meaning.

Reviews

Existing reviews of sign language processing are largely technical and out-of-date, written before the advent of deep learning. Most reviews focus on a specific subarea, such as the software and hardware used to recognize signs [2, 117, 61, 27]. Few reviews discuss multiple subfields of sign language technologies (e.g., recognition, translation, *and* generation). In this work, we provide a broader perspective that highlights common needs (e.g., datasets), and applications that blend multiple technologies. Unlike past reviews, we also articulate a call to action for the community, helping to prioritize problems facing the field.

Existing reviews also incorporate limited perspectives outside of computer science. In particular, few reviews incorporate the linguistic, social, and design perspectives needed to build sign language systems with real-world use. Some reviews consider a related discipline (e.g., linguistics in [27]), but do not consider the full spectrum of disciplines involved. This work integrates diverse interdisciplinary perspectives throughout, providing deeper insight into how technologies align with human experiences, the challenges facing the field, and opportunities for collaboration.

There is already interest among researchers in various fields in applying their methods to sign language applications. In particular, some technical reviews of gesture recognition touch on sign language recognition as an application domain (e.g., [120, 87, 108, 69]). These reviews focus on algorithms for detecting fingers, hands, and human gestures. However, by framing sign language recognition as an application area, they risk misrepresenting sign language recognition as a gesture recognition problem, ignoring the complexity of sign languages as well as the broader social context within which such systems must function. In this work, we provide linguistic and cultural context in conjunction with algorithms, to establish a more accurate representation of the space.

METHOD

To help answer our guiding questions, we convened a two-day workshop with leading experts in sign language processing and related fields. Many of these participants continued on to synthesize the workshop findings in this paper.

Participants

A total of 39 workshop attendees were recruited from universities and schools (18) and a tech company (21). Academic participants were based in departments spanning computer science, linguistics, education, psychology, and Deaf studies. Within computer science, specialists were present from computer vision, speech recognition, machine translation, machine learning, signal processing, natural language processing, computer graphics, human-computer interaction, and accessibility. Attendees from the tech company had roles in research, engineering, and program/product management. The number of participants present varied slightly over the two days.

Participants were demographically diverse:

- Nationality: Attendees were an international group, currently based in the U.S., Europe, and Asia.
- Experience: Career level ranged from recent college graduates through senior professors and executives.
- Gender: 25 male, 14 female
- Audiological status: 29 hearing (6 with Deaf immediate family members), 10 Deaf and hard of hearing (DHH)

Procedure

The workshop activities were structured to facilitate progress toward our three guiding questions (current landscape, biggest challenges, and calls to action). Day 1 provided the necessary domain-specific background knowledge, and Day 2 addressed our three questions as an interdisciplinary group. Interpreters and captioners were available throughout.

Day 1: Sharing interdisciplinary domain knowledge.

- **Domain Lectures:** A series of 45-minute talks, covering relevant domains and given by domain experts: Deaf culture, sign language linguistics, natural language processing, computer vision, computer graphics, and dataset curation.
- **Panel:** A 45-minute panel on Deaf users' experiences with, needs for, and concerns about technology, with a Deaf moderator and four Deaf panelists.

Day 2: Discussing problems and mapping the path forward.

- **Breakout Sessions:** Participants divided into smaller groups (8-9/group) to discuss specific topics, for 3.5 hours.

  The topic areas, outlined by the organizers and voted on by participants, were:

1. Sign Language Datasets
2. Sign Language Recognition & Computer Vision
3. Sign Language Modeling & NLP
4. Sign Language Avatars & Computer Graphics
5. Sign Language Technology User Interface/User Experience (UI/UX) Design

Each group focused on the following questions:

1. What is the state-of-the-art in this area?
2. What are the biggest current challenges in this area?
3. What are possible solutions to these challenges?
4. What is your vision of the future for this domain?
5. What is your call to action for the community?

- **Breakout Presentations:** Each breakout group reported back on their topic, through a slide presentation mixed with discussion with the larger group (about 20 minutes each).

In the following sections, we summarize the content generated through this workshop exercise, organized by our three guiding questions.

Q1: WHAT IS THE CURRENT LANDSCAPE?

In this section, we synthesize each group's formulation of the current state-of-the-art. We note that some topics overlap. In particular, data is central to progress on all fronts, so we start with a summary of the landscape in sign language datasets.

Datasets

Existing sign language datasets typically consist of videos of people signing. Video format can vary, and is often dependent on the recording device. For example, video cameras often produce MP4, OGG, or AVI format (among others). Motion-capture datasets have been curated, often by attaching sensors to a signer (e.g., [82, 52, 10]). These datasets can be pulled from to generate signing avatars, and are often curated for this purpose. Depth cameras can also be used to capture 3D positioning. For example, the Kinect includes a depth sensor and has been used to capture sign language data (e.g., [88, 28, 90]). Table 1 summarizes public sign language video corpora commonly used for sign language recognition. (See [77]'s survey for a more complete list of datasets, many of which are intended for linguistic research and education.)

The data collection method impacts both content and signer identity. For example, some corpora are formed of professional interpreters paid to interpret spoken content, such as news channels that provide interpreting [43, 74, 21]. Others are formed of expert signers paid to sign desired corpus content (e.g., [65, 124, 118]). Yet other corpora consist sign language videos posted on sites such as YouTube (e.g., [62]) – these posters may be fluent signers, interpreters, or sign language students; such videos are typically of "real-life" signs (i.e., self-generated rather than prompted). The geographic region where the data is collected also dictates which sign language is captured. Many datasets have been curated by researchers and startups in the U.S., where ASL is the primary language of the Deaf community, and consequently contain ASL. Fewer datasets have been curated of other sign languages, though some exist (e.g., [83]). The vocabulary size of these datasets varies from about 100–2,000 distinct signs (see Table 1).

Annotations may accompany video corpora. These annotations can demarcate components of signs (e.g., handshapes and movements), the identity and ordering of the signs, or a translation into another language like English. These annotations can take various formats, including linguistic notation systems (for sign components), English gloss (for sign identity and order), and English text (for translations). The annotations can be aligned at various levels of granularity. For example, the start and end of a handshape could be labeled, or the start and end of a full sentence. Generating annotations can be very time-intensive and expensive. Annotation software has been developed to facilitate annotating videos, and is often used by linguists studying the language (e.g., ELAN [112] and Anvil [70]). Because sign languages do not have a standard written form, large text corpora do not exist independent of videos.

Recognition & Computer Vision

Glove-based approaches to sign language recognition have been used to circumvent computer vision problems involved in recognizing signs from video. The first known work dates back to 1983, with a patent describing an electronic glove that recognized ASL fingerspellings based on a hardwired circuit [48]. Since then, many other systems have been built for "intrusive sign recognition," where signers are required to wear gloves, clothing, or other sensors to facilitate recognition (e.g., [23, 40, 81, 91, 27]).

Non-intrusive vision-based sign language recognition is the current dominant approach. Such systems minimize inconvenience to the signer (and, unlike gloves, have the potential to incorporate non-manual aspects of signing), but introduce complex computer vision problems. The first such work dates back to 1988, when Tamura et al. [110] built a system to recognize 10 isolated signs of Japanese Sign Language using skin color thresholding. As in that seminal work, many other systems focus on identifying individual signs (e.g., [49, 79]).

Real-world translation typically requires continuous sign language recognition [105, 41], where a continuous stream of signing is deciphered. Continuous recognition is a significantly more challenging and realistic problem than recognizing individual signs, confounded by epenthesis effects (insertion of extra features into signs), co-articulation (the ending of one sign affecting the start of the next), and spontaneous sign production (which may include slang, non-uniform speed, etc.).

To address the three-dimensionality of signs, some vision-based approaches use depth cameras [113, 123], multiple cameras [16] or triangulation for 3D reconstruction [99, 98]. Some use colored gloves to ease hand and finger tracking [26]. Recent advances in machine learning – i.e., deep learning and convolutional neural networks (CNNs) – have improved state-of-the-art computer vision approaches [76], though lack of sufficient training data currently limits the use of modern Artificial Intelligence (AI) techniques in this problem space.

Automatic recognition systems are transitioning from small, artificial vocabularies and tasks to larger real-world ones. Re-

Dataset	Vocabulary	Signers	Signer-independent	Videos	Continuous	Real-life
Purdue RVL-SLLL ASL [65]	104	14	no	2,576	yes	no
RWTH Boston 104 [124]	104	3	no	201	yes	no
Video-Based CSL [54]	178	50	no	25,000	yes	no
Signum [118]	465	(24 train, 1 test) - 25	yes	15,075	yes	no
MS-ASL [62]	1,000	(165 train, 37 dev, 20 test) - 222	yes	25,513	no	yes
RWTH Phoenix [43]	1,081	9	no	6,841	yes	yes
RWTH Phoenix SI5 [74]	1,081	(8 train, 1 test) - 9	yes	4,667	yes	yes
Devisign [22]	2,000	8	no	24,000	no	no

Table 1. Popular public corpora of sign language video. These datasets are commonly used for sign language recognition.

alistic scenarios are still very challenging for state-of-the-art algorithms. As such, recognition systems achieve only up to 42.8% letter accuracy [101] on a recently released real-life fingerspelling dataset. A real-life continuous sign language video dataset has also been released [43], and is used as a community benchmark. Given utterance- or sentence-level segmentation, recognition systems can reliably identify sign boundaries [72]. For such challenging datasets (still only covering a vocabulary of around 1,000 different signs), recognition algorithms can achieve a Word Error Rate (WER) of 22.9% [30] when trained and tested on the same signers, and a WER of 39.6% when trained and tested on different sets of signers.

Modeling & Natural Language Processing

Because sign languages are minority languages lacking data, the vast majority of work in MT and NLP focuses on spoken and written languages, not sign languages. While recognition handles the problem of identifying words or signs from complex signals (audio or video), MT and NLP typically address problems of processing language that has already been identified. These methods expect annotated data as input, which for spoken languages is commonly text (e.g., books, newspapers, or scraped text from the internet). Translation between spoken and signed languages (and vice versa) also typically requires intermediary representations of the languages that are computationally compatible.

Various notation systems are used for computational modeling. Human-generated annotations are often in gloss, a form of transliteration where written words in another language (e.g., English) are used to represents signs (e.g., ASL). Other writing systems have also been developed for people to use, including si5s [25] and SignWriting [109]. Separate notation systems have been developed for computers to represent sign languages during modeling and computation; HamNoSys [50] is one of the most popular, designed to capture detailed human movements and body positioning for computer modeling. To facilitate structured storage, XML-based markup languages have also been developed, e.g. Signing Gesture Markup Language (SiGML) [37], which is compatible with HamNoSys.

Sign language translation systems can either use predefined intermediary representations of the languages involved, or learn their own representations (which may not be human-understandable). Methods that use predefined representations are highly compatible with grammatical translation rules (e.g., [32, 125, 37, 116]). Methods that do not use such representations typically use some form of deep learning or neural networks, which learn model features (i.e. internal representations) that suit the problem. These methods have been used for recognition combined with translation, processing complete written sentences in parallel with signed sentences [76, 75, 21]. Such techniques are often used in computer vision systems, and overlap with works presented in the previous section.

Avatars & Computer Graphics

Sign language avatars (computer animations of humans) can provide content in sign language, making information accessible to DHH individuals who prefer sign language or have lower literacy in written language [56]. Because sign languages are not typically written, these videos can be preferable to text. Videos of human signers or artist-produced animations provide similar viewing experiences, but avatars are more appropriate when automatic generation is desirable (e.g., for a website with unstable content). Current pipelines typically generate avatars based on a symbolic representation of the signed content prepared by a human author (e.g. [3, 1, 12, 115, 7]). When the avatar is generated as part of a translation system (e.g., [66, 38]), an initial translation step converts spoken/written language into a symbolic representation of the sign language (as described in the previous section). Whether human-authored or automatically translated, a symbolic plan is needed for the sign-language message. While multiple representations have been proposed (e.g. [37, 12, 1]), there is no universal standard.

Beginning with this symbolic plan, pipelines generating avatars typically involve a series of complex steps (e.g., as outlined in [56, 46]). Animations for individual signs are often pulled from lexicons of individual signs. These motion-plans for the individual signs are produced in one of several ways: key-frame animations (e.g. [57]), symbolic encoding of sub-sign elements (e.g. [35]), or motion-capture recordings (e.g., [100, 46]). Similarly, non-manual signals are pulled from complementary datasets (e.g., [58]) or synthesized from models (e.g., [63]). These elements are combined to create an initial motion script of the content. Next, various parameters (e.g., speed, timing) are set by a human, set by a rule-based approach (e.g. [35]), or predicted via a trained machine-learning model (e.g., [3, 82]). Finally, computer animation software renders the animation based on this detailed movement plan.

The state-of-the-art in avatar generation is not fully automated; all parts of current pipelines currently require human intervention to generate smooth, coherent signing avatars. Prior research has measured the quality of avatar animations via perceptual and comprehension studies with DHH participants

[59], including methodological research [64] and shared resources for conducting evaluation studies [57].

UI/UX Design

The state-of-the-art of sign language output in user interfaces primarily centers around systems that use sign language video or animation content (e.g., computer-generated human avatars) to display information content. (Surveys of older work appear in [56] and [3].) These projects include systems that provide sign language animation content on webpages to supplement text content for users. There has also been some work on providing on-demand definitions of terminology in ASL (e.g., by linking to ASL dictionary resources [51]). As discussed in [51], prior work has found that displaying static images of signs provides limited benefit, and generally users have preferred interfaces that combine both text and sign content.

Research on designing interactive systems with sign language recognition technologies has primarily investigated how to create useful applications despite the limited accuracy and coverage of current technology for this task. This has often included research on tools for students learning ASL, either young children (e.g., [122]) or older students who are provided with feedback as to whether their signing is accurate [55]. While there have been various short-lived projects and specific industry efforts to create tools that can recognize full phrases of ASL to provide communication assistance, few systems are robust enough for real-world deployment or use.

Q2: WHAT ARE THE FIELD'S BIGGEST CHALLENGES?

In this section, we summarize the major challenges facing the field, identified by the interdisciplinary breakout groups.

Datasets

Public sign language datasets have shortcomings that limit the power and generalizability of systems trained on them.

Size: Modern, data-driven machine learning techniques work best in data-rich scenarios. Success in speech recognition, which in many ways is analogous to sign recognition, has been made possible by training on corpora containing millions of words. In contrast, sign language corpora, which are needed to fuel the development of sign language recognition, are several orders of magnitude smaller, typically containing fewer than 100,000 articulated signs. (See Table 2 for a comparison between speech and sign language datasets.)

Continuous Signing: Many existing sign language datasets contain individual signs. Isolated sign training data may be important for certain scenarios (i.e., creating a sign language dictionary), but most real-world use cases of sign language processing involve natural conversational with complete sentences and longer utterances.

Native Signers: Many datasets allow novices (i.e., students) to contribute, or contain data scraped from online sources (e.g., YouTube [62]) where signer provenance and skill is unknown. Professional interpreters, who are highly skilled but are often not native signers, are also used in many datasets (e.g., [42]). The act of interpreting also changes the execution (e.g., by simplifying the style and vocabulary, or signing slower for understandability). Datasets of native signers are needed to build models that reflect this core user group.

Signer Variety: The small size of current signing datasets and over-reliance on content from interpreters mean that current datasets typically lack signer variety. To accurately reflect the signing population and realistic recognition scenarios, datasets should include signers that vary by: gender, age, clothing, geography, culture, skin tone, body proportions, disability, fluency, background scenery, lighting conditions, camera quality, and camera angles. It is also crucial to have signer-independent datasets, which allow people to assess generalizability by training and testing on different signers. Datasets must also be generated for different sign languages (i.e., in addition to ASL).

Recognition & Computer Vision

Despite the large improvements in recent years, there are still many important and unsolved recognition problems, which hinder real-world applicability.

Depiction: Depiction refers to visually representing or enacting content in sign languages (see Background & Related Work), and poses unique challenges for recognition and translation. Understanding depiction requires exposure to Deaf culture and linguistics, which the communities driving progress in computer vision generally lack. Sign recognition algorithms are often based on speech recognition, which does not handle depictions (which are uncommon and unimportant in speech). As a result, current techniques cannot handle depictions. It is also difficult to create depiction annotations. Countless depictions can express the same concept, and annotation systems do not have a standard way to encode this richness.

Annotations: Producing sign language annotations, the machine-readable inputs needed for supervised training of AI models, is time consuming and error prone. There is no standardized annotation system or level of annotation granularity. As a result, researchers are prevented from combining annotated datasets to increase power, and must handle low inter-annotator agreement. Annotators must also be trained extensively to reach sufficient proficiency in the desired annotation system. Training is expensive, and constrains the set of people who can provide annotations beyond the already restricted set of fluent signers. The lack of a standard written form also prevents learning from naturally generated text – e.g., NLP methods that expect text input, using parallel text corpora to learn corresponding grammar and vocabulary, and more generally leveraging ubiquitous text resources.

Generalization: Generalization to unseen situations and individuals is a major difficulty of machine learning, and sign language recognition is no exception. Larger, more diverse datasets are essential for training generalizable models. We outlined key characteristics of such datasets in the prior section on dataset challenges. However, generating such datasets can be extremely time-consuming and expensive.

Modeling & Natural Language Processing

The main challenge facing modeling and NLP is the inability to apply powerful methods used for spoken/written languages due to language structure differences and lack of annotations.

	Sign Language	Speech
Modality	visual-gestural	aural-oral
Articulators	manual, non-manual	vocal tract
Seriality	low	high
Simultaneity	high	low
Iconicity	high	low
Task	recognition, generation, translation	recognition, generation, translation
Typical articulated corpus size	<100,000 signs	5 million words
Typical annotated corpus size	<100,000 signs	1 billion words
Typical corpus vocabulary size	1,500 signs	300,000 words
What is being modelled	1,500 whole signs	1,500 tri-phonemes
Typical corpus number of speakers	10	1,000

Table 2. Comparison of sign language vs. speech datasets. Existing sign language corpora are orders of magnitude smaller than speech corpora. Because sign languages are not typically written, parallel written corpora do not exist for sign languages, as they do for spoken (and written) languages.

Structural Complexity: Many MT and NLP methods were developed for spoken/written languages. However, sign languages have a number of structural differences from these languages. These differences mean that straightforward application of MT and NLP methods will fail to capture some aspects of sign languages or simply not work. In particular, many methods assume that one word or concept is executed at a time. However, many sign languages are multi-channel, for instance conveying an object and its description simultaneously. Many methods also assume that context does not change the word being uttered; however, in sign languages, content can be spatially organized and interpretation directly dependent on that spatial context.

Annotations: Lack of reliable, large-scale annotations are a barrier to applying powerful MT and NLP methods to sign languages. These methods typically take annotations as input, commonly text. Because sign languages do not have a standard written form or a standard annotation form, we do not have large-scale annotations to feed these methods. Lack of large-scale annotated data is similarly a problem for training recognition systems, as described in the previous section.

Avatars & Computer Graphics

Avatar generation faces a number of technical challenges in creating avatars that are acceptable to Deaf users (i.e., pleasing to view, easy to understand, representative of the Deaf community, etc.). Some of these problems may be addressed by including Deaf people in the generation process [71].

Uncanny Valley: Sign language avatars are subject to an uncanny valley [86]. Avatars that are either very cartoonish or very human-like are fairly pleasing, but in-between can be disconcerting. For example, in addition to providing semantically meaningful non-manual cues (e.g., raised eyebrows indicating a question), avatars must also have varied, natural facial expressions (i.e., not a robotic, stoic expression throughout). It can be difficult to design avatars that fall outside of this valley.

Realistic Transitions: To illustrate why transitions between signs are difficult, consider a generation system that pulls from motion-capture libraries. The system can pull complete sign executions, but must then piece together these executions. One sign might end with the hands in one position, while the subsequent sign starts with the hands in another position, and the software must create a smooth, natural transition between.

Modeling Modulations: In some sign languages, adjectives and adverbs are executed by modulating a noun or verb. For example, in ASL, PLANE RIDE is executed by moving a certain handshape through the air. BUMPY PLANE RIDE is identical, but with the movement made bumpy. Infinite such descriptors can be executed, and capturing them all in a motion-capture database is infeasible. Acceptable abstractions have not been standardized (e.g., in a writing system), so it is unclear how much real-life variation avatars must portray.

Finding Model Holes: It is difficult to find holes in generation models, because the language space is large and rich, and the number of ways that signs can be combined in sequence grows exponentially. Testing all grammatical structures empirically is not scalable. This "unknown unknown" problem is common to other machine learning areas (e.g., speech recognition [53]).

Public Motion-Capture Datasets Many motion-capture datasets used for avatar generation are owned by particular companies or research groups. Because they are not publicly available, research in this area can be impeded.

UI/UX Design

Sign language UI/UX design is currently confounded by technical limitations that require carefully scoped projects, many potential use cases requiring different solutions, and design choices that may have powerful social ramifications.

Technical Limitations A long-term goal in this space is full universal design of conversational agents. For example, if a system supports speech-based or text chat interaction, then it should also support input and output in sign language. However, given the current limitations of the component technologies, it may be useful for researchers to focus on more near-term research aims: for instance, if we have a sign language recognition system capable of recognizing some finite number of signs or phrases, then what types of applications can be supported within this limit (for different vocabulary sizes)?

Varied Use Cases: There are a huge number of use cases for sign language processing, requiring different interface designs. For example, sign language recognition could be useful for placing a meal order in a drive-through restaurant, or for commanding a personal assistant. Similarly, sign language generation may be used in various situations. For people who want to create websites that present signed content, avatars may be the most reasonable solution, as they allow for ease

in editability, creation from anywhere, and scalability (cost). However, people also want websites to be searchable and indexable, and videos and animations are difficult for current text-based search engines to index and search. Combining text and video introduces layout problems, especially when text is automatically replaced with video. These situations, and many others, have drastically different design criteria.

Language and Dialect Choice: Many different sign languages exist, with many dialects for each. Choosing which one(s) a system will recognize or portray is a difficult problem with societal implications. Minorities within Deaf communities may be further marginalized if their dialects are not represented. Similarly, failure to represent other types of diversity – e.g., gender, race, education level, etc. – could also be detrimental.

Q3: WHAT ARE THE CALLS TO ACTION?

In this section, we outline an interdisciplinary call to action for the research community working on any piece of the end-to-end sign language processing pipeline. Once stated, these calls to action may seem intuitive, but have not previously been articulated, and have until now been largely disregarded.

Deaf Involvement

In developing sign language processing, Deaf community involvement is essential at all levels, in order to design systems that actually match user needs, are usable, and to facilitate adoption of the technology. An all-hearing team lacks the lived experience of Deafness, and is removed from the use cases and contexts within which sign language software must function. Even hearing people with strong ties to the Deaf community are not in a position to speak for Deaf needs. Additionally, because of their perceived expertise in Deaf matters, they are especially susceptible to being involved in Deaf-hearing power imbalances. People who do not know a sign language also typically make incorrect assumptions about sign languages – e.g., assuming that a particular gesture always translates to a particular spoken/written word. As a result, all-hearing teams are ill-equipped to design software that will be truly useful.

It is also important to recognize individual and community freedoms in adopting technology. Pushing a technology can lead to community resentment, as in the case of cochlear implants for many members of sign language communities [104]. Disrespecting the Deaf community's ownership over sign languages also furthers a history of audism and exclusion, which can result in the Deaf community rejecting the technology. For these reasons, a number of systems built by hearing teams to serve the Deaf community have failed or receive mixed reception (e.g., sign language gloves [39]).

Deaf contributors are essential at *every step* of research and development. For example, involvement in the creation, evaluation, and ownership of sign language datasets is paramount to creating high-quality data that accurately represents the community, can address meaningful problems, and avoids cultural appropriation. Future datasets might take cultural competency into account by 1) being open-source and publicly available, 2) providing cultural context for challenges to ensure that computer vision experts competing on algorithmic performance understand the nature, complexity, and history of sign languages, and/or 3) providing more appropriate metrics developed by the Deaf community, beyond the current standard of WER. Similarly, Deaf community involvement is fundamental to the creation of appropriate computational models, interface design, and overall systems.

The importance of Deaf involvement is heightened by technology's impact on language. Written English is evolving right now with new spellings based on technological constraints like character limits on Twitter, and difficulty typing long phrases on phone keyboards. It is possible that signers would similarly adapt sign languages to better suit the constraints of computing technologies. For example, people might simplify vocabulary to aid recognition software, constrict range of motion to fit the technical limits of video communications [68], or abstract away richness to support standardized writing or annotation.

Call 1: Involve Deaf team members throughout.

Deaf involvement and leadership are crucial for designing systems that are useful to users, respecting Deaf ownership of sign languages, and securing adoption.

Application Domain

There are many different application domains for sign language processing. Situations where an interpreter would be beneficial but is not available are one class of applications. This includes any point of sale, restaurant service, and daily spontaneous interactions (for instance with a landlord, colleagues, or strangers). Developing personal assistant technologies that can respond to sign language is another compelling application area. Each of these scenarios requires different solutions. Furthermore, these different use cases impose unique constraints on every part of the pipeline, including the content, format, and size of training data, the properties of algorithms, as well as the interface design. Successful systems require buy-in from the Deaf community, so ensuring that solutions handle application domains appropriately is essential.

Technical limitations impact which domains are appropriate to tackle in the near-term, and inform intermediary goals that which will ultimately inform end-to-end systems. Many of these intermediary goals are worth pursuing in and of themselves, and offer bootstrapping benefits toward longer-term goals. For example, a comprehensive, accurate sign language dictionary that lets users look up individual signs would be an important resource for sign language users and learners alike, and would also inform model design for continuous sign language recognition. In addition, support for everyday use of sign language writing would make text-based resources accessible to sign language users in their language of choice, and would also organically generate an annotated corpus of sign language that could be used to learn language structure.

Call 2: Focus on real-world applications.

Sign language processing is appropriate for specific domains, and the technology has limitations. Datasets, algorithms, interfaces, and overall systems should be built to serve real-world use cases, and account for real-world constraints.

Interface Design

The field currently lacks fundamental research on how users interact with sign language technology. A number of systems have been developed explicitly serving sign language users (e.g., messaging services [111, 121], games [80, 15], educational tools [94, 4], webpages [92], dictionaries [102, 14], and writing support [11, 13]). However, accompanying user studies typically focus on evaluating a single system, and do not outline principles of interaction that apply across systems. As a result, each team developing a new system must design their interface largely from scratch, uninformed by general design guidelines based on research.

Since many technologies required for end-to-end sign language translation are under development, it may be necessary for researchers to use Wizard-of-Oz style testing procedures (e.g., [31]) to better understand how Deaf users would react to various types of user-interface designs. Recent work has used such approaches. For instance, researchers have used Wizard-of-Oz methodologies to study how Deaf users would like to issue commands to personal assistants [95] or how Deaf users may benefit from a tool that enables ASL dictionary lookup on-demand when reading English text webpages [51].

Returning to the personal assistant application mentioned above, a Wizard-of-Oz methodology could be used to investigate interaction questions, such as how the user might "wake up" the system so it expects a command, and how the system might visually acknowledge a signed command (e.g., by presenting written-language text onscreen) and provide a response to the user (e.g., as written-language text or as sign-language animation). Additionally, such simulations may also be used to determine how good these technologies must be before they are acceptable to users, i.e., what threshold of recognition accuracy is acceptable to users in specific use cases. Such work can set an agenda for researchers investigating the development of core sign-language recognition or synthesis technologies.

Call 3: Develop user-interface guidelines for sign language systems.

Because sign language processing is still developing, we lack a systematic understanding of how people interact with it. Guidelines and error metrics for effective system design would support the creation of consistently effective interfaces.

Datasets

As highlighted throughout this work, few large-scale, publicly available sign language corpora exist. Moreover, the largest public datasets are orders of magnitude smaller than those of comparable fields like speech recognition. The lack of large-scale public datasets shifts the focus from algorithmic and system development to data curation. Establishing large, appropriate corpora would expedite technical innovation.

In particular, the field would benefit from a larger body of research involving reproducible tasks. Publicly available data and competitive evaluations are needed to create interest, direct research towards the challenges that matter (tackling depiction, generalizing to unseen signers, real-life data), and increase momentum. Furthermore, having open-source implementations of full pipelines would also foster faster adoption.

There are four main approaches for collecting signing data, each of which has strengths and weaknesses. Developing multiple public data resources that span these four approaches may be necessary in order to balance these tradeoffs.

1. Scraping video sites (e.g., YouTube) has many potential benefits: low cost, rapid collection of many videos, the naturalistic nature of the data, and potential diversity of participants. Its pitfalls include: privacy and consent of people in the videos, variability in signing quality, and lack of accompanying annotations.

2. Crowdsourcing data through existing platforms (e.g., Amazon Mechanical Turk) or customized sites (e.g., [14]) offers potential cost savings (particularly if participants contribute data for free), and the ability to reach diverse contributors (i.e., by removing geographic constraints). However, crowdsourcing is subject to quality control issues. In paid systems people may rush or "cheat" to earn more money, and in unpaid learning activities, well-intentioned learners may submit low-quality or incorrect data.

3. Bootstrapping, where products are released with limitations and gather data during use, is common to other AI domains (e.g., voice recognition [97]). This approach is cheap, collects highly naturalistic data, and may scale well. However, privacy and informed consent are potential pitfalls, and there is a cold-start problem – can a useful application be created from current datasets to support this bootstrapping process, and can it achieve a critical mass of users?

4. In-lab collection allows for customized, high-end equipment such as high-resolution, high-frame-rate cameras, multiple cameras, depth-cameras, and motion-capture suits. However, this type of controlled collection may result in less naturalistic content, higher costs that limit scalability, and lower participant diversity due to geographic constraints. Models trained on such high-quality data also may not generalize to users with low-quality phone or laptop cameras.

Some metadata impacting data utility can only be gathered at the time of capture. In particular, demographics may be important for understanding biases and generalizability of systems trained on the data [44]. Key demographics include signing fluency, language acquisition age, education (level, Deaf vs. mainstream), audiological status, socioeconomic status, gender, race/ethnicity, and geography. Such metadata can also benefit linguistics, Deaf studies, and other disciplines.

Metadata regarding the data collection process itself (i.e., details enabling replication) are also vital to include so that others can add to the dataset. For example, if a dataset is gathered in the U.S., researchers in other countries could replicate the collection method to increase geographic diversity.

Call 4: Create larger, more representative, public video datasets.

Large datasets with diverse signers are essential for training software to perform well for diverse users. Public availability is important for spurring developments, and for ensuring that the Deaf community has equal ownership.

Annotations

A standard annotation system would expedite development of sign language processing. Datasets annotated with the standard system could easily be combined and shared. Software systems built to be compatible with that annotation system would then have much more training data at their disposal. A standard system would also reduce annotation cost and errors. As described earlier, the lack of standardization results in expensive training (and re-training) of annotators, and ambiguous, error-prone annotations.

Designing the annotation system to be appropriate for everyday reading and writing, or developing a separate standard writing system, would provide addition benefits. With such a system, email clients, text editors, and search engines would become newly usable in sign languages without translating into a spoken/written language. As they write, users would also produce a large annotated sign language corpus of naturally generated content, which could be used to better train models. However, establishing a standard writing system requires the Deaf community to reach consensus on how much of the live language may be abstracted away. Any writing system loses some of the live language (i.e., a transcript of a live speech in English loses pitch, speed, intonation, and emotional expression). Sign languages will be no different.

Computer-aided annotation software has been proposed (e.g., [29, 33, 24]), but could provide increased support due to recent advances in deep learning applied to sign language recognition. Current sign language modeling techniques could be used to aid the annotation process in terms of both segmenting and transcribing the input video. Aided annotation should leverage advances in modeling whole signs and also sign subunits [73, 72]. Annotation support tools could also alleviate problems with annotating depictions, as they could propose annotations conditioned on the translation and hence circumvent the problem of detailing the iconic nature of these concepts.

Call 5: Standardize the annotation system and develop software for annotation support.

Annotations are essential to training recognition systems, providing inputs to NLP and MT software, and generating signing avatars. Standardization would support data sharing, expand software compatibility, and help control quality. Annotation support would help improve accuracy, reliability, and cost.

CONTRIBUTIONS

This paper provides an interdisciplinary perspective on the field of sign language processing. For computer scientists and technologists, it provides key background on Deaf culture and sign language linguistics that is often lacking, and contextualizes relevant subdomains they may be working within (i.e., HCI, computer vision, computer graphics, MT, NLP). For readers outside of computer science, it provides an overview of how sign language processing works, and helps to explain the challenges that current technologies face.

In synthesizing the state-of-the-art from an interdisciplinary perspective, this paper provides orientation for researchers in any domain, in particular those entering the field. Unlike disciplinary reviews that focus on relevant work in a particular domain, we relate these domains to one another, and show how sign language processing is dependent on all of them.

In summarizing the current challenges, this work highlights opportunities for interdisciplinary collaboration. Many of the problems facing the field cross disciplines. In particular, questions of how to create datasets, algorithms, user interfaces, and a standard annotation system that meet technical requirements, reflect linguistics of the language, and are accepted by the Deaf community are large, open problems that will require strong, interdisciplinary teams.

Finally, in articulating a call to action, this work helps researchers prioritize efforts to focus on the most pressing and important problems. Lack of data (in particular large, annotated, representative, public datasets) is arguably the biggest obstacle currently facing the field. This problem is confounded by the relatively small pool of potential contributors, recording requirements, and lack of standardized annotations. Because data collection is difficult and costly, companies and research groups are also incentivised to keep data proprietary. Without sufficient data, system performance will be limited and unlikely to meet the Deaf community's standards.

Our workshop methodology used to provide this interdisciplinary perspective on sign language processing can be used as a model for other similarly siloed fields. While the general structure of the workshop is directly duplicable, some work would need to be done to tailor it to other fields (e.g., identifying relevant domains and domain experts).

CONCLUSION

In this paper, we provide an interdisciplinary overview of sign language recognition, generation, and translation. Past work on sign language processing has largely been conducted by experts in different domains separately, limiting real-world utility. In this work, we assess the field from an interdisciplinary perspective, tackling three questions: 1) What does an interdisciplinary view of the current landscape reveal? 2) What are the biggest challenges facing the field? and 3) What are the calls to action for people working in the field?

To address these questions, we ran an interdisciplinary workshop with 39 domain experts with diverse backgrounds. This paper presents the interdisciplinary workshop's findings, providing key background for computer scientists on Deaf culture and sign language linguistics that is often overlooked, a review of the state-of-the-art, a set of pressing challenges, and a call to action for the research community. In doing so, this paper serves to orient readers both within and outside of computer science to the field, highlights opportunities for interdisciplinary collaborations, and helps the research community prioritize which problems to tackle next (data, data, data!).

ACKNOWLEDGMENTS

The authors would like to thank all workshop participants. We also thank Bill Thies for helpful discussions and prudent advice. This material is based on work supported by Microsoft, NIH R21-DC016104, and NSF awards 1462280, 1746056, 1749376, 1763569, 1822747, 1625793, and 1749384.

REFERENCES

1. Nicoletta Adamo-Villani and Ronnie B. Wilbur. 2015. ASL-Pro: American Sign Language Animation with Prosodic Elements. In *Universal Access in Human-Computer Interaction. Access to Interaction*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, 307–318.
2. M Ebrahim Al-Ahdal and Md Tahir Nooritawati. 2012. Review in sign language recognition systems. In *2012 IEEE Symposium on Computers & Informatics (ISCI)*. IEEE, 52–57.
3. Sedeeq Al-khazraji, Larwan Berke, Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2018. Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 259–270.
4. Anwar AlShammari, Asmaa Alsumait, and Maha Faisal. 2018. Building an Interactive E-Learning Tool for Deaf Children: Interaction Design Process Framework. In *2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*. IEEE, 85–90.
5. UN General Assembly. 2006. Convention on the Rights of Persons with Disabilities. *GA Res* 61 (2006), 106.
6. British Deaf Association. 2015. George Veditz Quote - 1913. (2015). `https://vimeo.com/132549587` Accessed 2019-04-22.
7. J Andrew Bangham, SJ Cox, Ralph Elliott, JRW Glauert, Ian Marshall, Sanja Rankov, and Mark Wells. 2000. Virtual signing: Capture, animation, storage and transmission-an overview of the visicast project. (2000).
8. H-Dirksen L Bauman. 2004. Audism: Exploring the Metaphysics of Oppression. *Journal of deaf studies and deaf education* 9, 2 (2004), 239–246.
9. H-Dirksen L Bauman and Joseph J Murray. 2014. *Deaf Gain: Raising the Stakes for Human Diversity*. U of Minnesota Press.
10. Bastien Berret, Annelies Braffort, and others. 2016. Collecting and Analysing a Motion-Capture Corpus of French Sign Language. In *Workshop on the Representation and Processing of Sign Languages*.
11. Claudia Savina Bianchini, Fabrizio Borgia, Paolo Bottoni, and Maria De Marsico. 2012. SWift: a SignWriting improved fast transcriber. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 390–393.
12. Annelies Braffort, Michael Filhol, Maxime Delorme, Laurence Bolot, Annick Choisier, and Cyril Verrecchia. 2016. KAZOO: A Sign Language Generation Platform Based on Production Rules. *Univers. Access Inf. Soc.* 15, 4 (Nov. 2016), 541–550. `DOI: http://dx.doi.org/10.1007/s10209-015-0415-2`
13. Danielle Bragg, Raja Kushalnagar, and Richard Ladner. 2018. Designing an Animated Character System for American Sign Language. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 282–294.
14. Danielle Bragg, Kyle Rector, and Richard E Ladner. 2015. A User-Powered American Sign Language Dictionary. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1837–1848.
15. Helene Brashear, Valerie Henderson, Kwang-Hyun Park, Harley Hamilton, Seungyon Lee, and Thad Starner. 2006. American sign language recognition in game development for deaf children. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 79–86.
16. Helene Brashear, Thad Starner, Paul Lukowicz, and Holger Junker. 2003. Using Multiple Sensors for Mobile Sign Language Recognition. In *7th IEEE International Symposium on Wearable Computers*. IEEE.
17. Diane Brentari. 1996. Trilled Movement: Phonetic Realization and Formal Representation. *Lingua* 98, 1-3 (1996), 43–71.
18. Diane Brentari. 2011. Handshape in Sign Language Phonology. *Companion to phonology* (2011), 195–222.
19. Diane Brentari. 2018. Representing Handshapes in Sign Languages Using Morphological Templates1. *Gebärdensprachen: Struktur, Erwerb, Verwendung* 13 (2018), 145.
20. Diane Brentari, Jordan Fenlon, and Kearsy Cormier. 2018. Sign Language Phonology. (2018). `DOI:http://dx.doi.org/10.1093/acrefore/9780199384655.013.117`
21. Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural Sign Language Translation. In *IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT.
22. Xiujuan Chai, Hanjie Wang, and Xilin Chen. 2014. *The DEVISIGN Large Vocabulary of Chinese Sign Language Database and Baseline Evaluations*. Technical Report. Key Lab of Intelligent Information Processing of Chinese Academy of Sciences. 00000.
23. C. Charayaphan and A. E. Marble. 1992. Image Processing System for Interpreting Motion in American Sign Language. *Journal of Biomedical Engineering* 14, 5 (Sept. 1992), 419–425. `DOI: http://dx.doi.org/10.1016/0141-5425(92)90088-3`
24. Émilie Chételat-Pelé, Annelies Braffort, and J Véronis. 2008. Sign Language Corpus Annotation: toward a new Methodology.. In *LREC*.
25. Adrean Clark. 2012. *How to Write American Sign Language*. ASLwrite.
26. Helen Cooper and Richard Bowden. 2010. Sign Language Recognition Using Linguistically Derived Sub-Units.... In *LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*. Valetta, Malta, 57–61.

27. Helen Cooper, Brian Holt, and Richard Bowden. 2011. Sign Language Recognition. In *Visual Analysis of Humans*. Springer, 539–562.

28. Helen Cooper, Eng-Jon Ong, Nicolas Pugeault, and Richard Bowden. 2012. Sign Language Recognition Using Sub-Units. *The Journal of Machine Learning Research* 13, 1 (2012), 2205–2231.

29. Onno A. Crasborn. 2015. Transcription and Notation Methods. In *Research Methods in Sign Language Studies*. John Wiley & Sons, Ltd, 74–88. DOI: http://dx.doi.org/10.1002/9781118346013.ch5

30. R. Cui, H. Liu, and C. Zhang. 2019. A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training. *IEEE Transactions on Multimedia* 0 (2019), 1–1. DOI: http://dx.doi.org/10.1109/TMM.2018.2889563

31. Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studiesâĂŤwhy and how. *Knowledge-based systems* 6, 4 (1993), 258–266.

32. Maksym Davydov and Olga Lozynska. 2017. Information system for translation into Ukrainian sign language on mobile devices. In *2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Vol. 1. IEEE, 48–51.

33. Philippe Dreuw and Hermann Ney. 2008. Towards automatic sign language annotation for the elan tool. In *Workshop Programme*. 50.

34. Paul Dudis. 2004. Depiction of events in ASL: Conceptual integration of temporal components. (2004).

35. Sarah Ebling and John Glauert. 2016. Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community. *Universal Access in the Information Society* 15, 4 (01 Nov 2016), 577–587. DOI: http://dx.doi.org/10.1007/s10209-015-0408-1

36. Richard Clark Eckert and Amy June Rowley. 2013. Audism: A Theory and Practice of Audiocentric Privilege. *Humanity & Society* 37, 2 (2013), 101–130.

37. Ralph Elliott, John RW Glauert, JR Kennaway, and Ian Marshall. 2000. The development of language processing support for the ViSiCAST project. In *ASSETS*, Vol. 2000. 4th.

38. Ralph Elliott, John RW Glauert, JR Kennaway, Ian Marshall, and Eva Safar. 2008. Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society* 6, 4 (2008), 375–391.

39. Michael Erard. 2017. Why Sign-Language Gloves Don't Help Deaf People. *The Atlantic* 9 (2017). https://www.theatlantic.com/technology/archive/2017/11/why-sign-language-gloves-dont-help-deaf-people/545441/

40. S. S. Fels and G. E. Hinton. 1993. Glove-Talk: A Neural Network Interface between a Data-Glove and a Speech Synthesizer. *IEEE Transactions on Neural Networks* 4, 1 (Jan. 1993), 2–8. DOI: http://dx.doi.org/10.1109/72.182690

41. Jens Forster, Christian Oberdörfer, Oscar Koller, and Hermann Ney. 2013. Modality Combination Techniques for Continuous Sign Language Recognition. In *Iberian Conference on Pattern Recognition and Image Analysis (Lecture Notes in Computer Science 7887)*. Springer, Madeira, Portugal, 89–99.

42. Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater, and Hermann Ney. 2012. RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. In *International Conference on Language Resources and Evaluation*. Istanbul, Turkey, 3785–3789.

43. 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 *International Conference on Language Resources and Evaluation*. Reykjavik, Island, 1911–1916.

44. Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford. 2018. Datasheets for datasets. *arXiv preprint arXiv:1803.09010* (2018).

45. Ann E Geers, Christine M Mitchell, Andrea Warner-Czyz, Nae-Yuh Wang, Laurie S Eisenberg, CDaCI Investigative Team, and others. 2017. Early sign language exposure and cochlear implantation benefits. *Pediatrics* 140, 1 (2017).

46. Sylvie Gibet, Nicolas Courty, Kyle Duarte, and Thibaut Le Naour. 2011. The SignCom system for data-driven animation of interactive virtual signers: Methodology and Evaluation. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 1, 1 (2011), 6.

47. Neil Stephen Glickman. 1993. Deaf Identity Development: Construction and Validation of a Theoretical Model. (1993).

48. Gary J. Grimes. 1983. Digital Data Entry Glove Interface Device. (Nov. 1983). US Patent.

49. Kirsti Grobel and Marcell Assan. 1997. Isolated sign language recognition using hidden Markov models. In *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, Vol. 1. IEEE, 162–167.

50. Thomas Hanke. 2004. HamNoSys-representing sign language data in language resources and language processing contexts. In *LREC*, Vol. 4. 1–6.

51. Dhananjai Hariharan, Sedeeq Al-khazraji, and Matt Huenerfauth. 2018. Evaluation of an English Word Look-Up Tool for Web-Browsing with Sign Language Video for Deaf Readers. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 205–215.

52. Alexis Heloir, Sylvie Gibet, Franck Multon, and Nicolas Courty. 2005. Captured Motion Data Processing for Real Time Synthesis of Sign Language. In *International Gesture Workshop*. Springer, 168–171.

53. Hynek Hermansky. 2013. Multistream recognition of speech: Dealing with unknown unknowns. *Proc. IEEE* 101, 5 (2013), 1076–1088.

54. Jie Huang, Wengang Zhou, Qilin Zhang, Houqiang Li, and Weiping Li. 2018. Video-Based Sign Language Recognition without Temporal Segmentation. In *Proceedings of the Thirty-Second {AAAI} Conference on Artificial Intelligence*. New Orleans, Louisiana, USA.

55. Matt Huenerfauth, Elaine Gale, Brian Penly, Sree Pillutla, Mackenzie Willard, and Dhananjai Hariharan. 2017. Evaluation of language feedback methods for student videos of american sign language. *ACM Transactions on Accessible Computing (TACCESS)* 10, 1 (2017), 2.

56. Matt Huenerfauth and V Hanson. 2009. Sign language in the interface: access for deaf signers. *Universal Access Handbook. NJ: Erlbaum* 38 (2009).

57. Matt Huenerfauth and Hernisa Kacorri. 2014. Release of Experimental Stimuli and Questions for Evaluating Facial Expressions in Animations of American Sign Language. In *Proceedings of the the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel, The 9th International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland.*

58. Matt Huenerfauth, Mitch Marcus, and Martha Palmer. 2006. *Generating American Sign Language classifier predicates for English-to-ASL machine translation*. Ph.D. Dissertation. University of Pennsylvania.

59. Matt Huenerfauth, Liming Zhao, Erdan Gu, and Jan Allbeck. 2007. Evaluating American Sign Language Generation Through the Participation of Native ASL Signers. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets '07)*. ACM, New York, NY, USA, 211–218. DOI: http://dx.doi.org/10.1145/1296843.1296879

60. Tom Humphries. 1975. Audism: The Making of a Word. *Unpublished essay* (1975).

61. Saba Joudaki, Dzulkifli bin Mohamad, Tanzila Saba, Amjad Rehman, Mznah Al-Rodhaan, and Abdullah Al-Dhelaan. 2014. Vision-based sign language classification: a directional review. *IETE Technical Review* 31, 5 (2014), 383–391.

62. Hamid Reza Vaezi Joze and Oscar Koller. 2018. MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language. *arXiv:1812.01053 [cs]* (Dec. 2018).

63. Hernisa Kacorri and Matt Huenerfauth. 2016. Continuous Profile Models in ASL Syntactic Facial Expression Synthesis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 2084–2093. DOI: http://dx.doi.org/10.18653/v1/P16-1196

64. Hernisa Kacorri, Matt Huenerfauth, Sarah Ebling, Kasmira Patel, Kellie Menzies, and Mackenzie Willard. 2017. Regression Analysis of Demographic and Technology-Experience Factors Influencing Acceptance of Sign Language Animation. *ACM Trans. Access. Comput.* 10, 1, Article 3 (April 2017), 33 pages. DOI: http://dx.doi.org/10.1145/3046787

65. Avi C. Kak. 2002. Purdue RVL-SLLL ASL Database for Automatic Recognition of American Sign Language. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces (ICMI '02)*. IEEE Computer Society, Washington, DC, USA, 167–172. DOI: http://dx.doi.org/10.1109/ICMI.2002.1166987

66. Kostas Karpouzis, George Caridakis, S-E Fotinea, and Eleni Efthimiou. 2007. Educational resources and implementation of a Greek sign language synthesis architecture. *Computers & Education* 49, 1 (2007), 54–74.

67. Jonathan Keane, Diane Brentari, and Jason Riggle. 2012. Coarticulation in ASL Fingerspelling. In *Proceedings of the North East Linguistic Society*, Vol. 42.

68. Elizabeth Keating, Terra Edwards, and Gene Mirus. 2018. Cybersign and new Proximities: Impacts of New Communication Technologies on Space and Language. *Journal of Pragmatics* 40, 6 (2018), 1067–1081.

69. Rafiqul Zaman Khan and Noor Adnan Ibraheem. 2012. Hand Gesture Recognition: a Literature Review. *International journal of artificial Intelligence & Applications* 3, 4 (2012), 161.

70. Michael Kipp. 2017. Anvil. (2017). https://www.anvil-software.org/ Accessed 2019-04-29.

71. Michael Kipp, Quan Nguyen, Alexis Heloir, and Silke Matthes. 2011. Assessing the Deaf User Perspective on Sign Language Avatars. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 107–114.

72. Oscar Koller, Necati Cihan Camgoz, Hermann Ney, and Richard Bowden. 2019. Weakly Supervised Learning with Multi-Stream CNN-LSTM-HMMs to Discover Sequential Parallelism in Sign Language Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 0 (2019), 15.

73. Oscar Koller, Hermann Ney, and Richard Bowden. 2016. Deep Hand: How to Train a CNN on 1 Million Hand Images When Your Data Is Continuous and Weakly Labelled. In *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA, 3793–3802.

74. Oscar Koller, Sepehr Zargaran, and Hermann Ney. 2017. Re-Sign: Re-Aligned End-To-End Sequence Modelling With Deep Recurrent CNN-HMMs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA, 4297–4305.

75. Oscar Koller, Sepehr Zargaran, Hermann Ney, and Richard Bowden. 2016. Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. In *British Machine Vision Conference*. York, UK.

76. Oscar Koller, Sepehr Zargaran, Hermann Ney, and Richard Bowden. 2018. Deep Sign: Enabling Robust Statistical Continuous Sign Language Recognition via Hybrid CNN-HMMs. *International Journal of Computer Vision* 126, 12 (Dec. 2018), 1311–1325. DOI: http://dx.doi.org/10.1007/s11263-018-1121-3

77. Reiner Konrad. 2012. Sign language corpora survey. (2012). https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/SL-Corpora-Survey_update_2012.pdf

78. Harlan Lane. 2017. A Chronology of the Oppression of Sign Language in France and the United States. In *Recent perspectives on American Sign Language*. Psychology Press, 119–161.

79. Simon Lang, Marco Block, and Raúl Rojas. 2012. Sign language recognition using kinect. In *International Conference on Artificial Intelligence and Soft Computing*. Springer, 394–402.

80. Seungyon Lee, Valerie Henderson, Harley Hamilton, Thad Starner, Helene Brashear, and Steven Hamilton. 2005. A gesture-based American Sign Language Game for Deaf Children. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1589–1592.

81. Rung-Huei Liang and Ming Ouhyoung. 1998. A Real-Time Continuous Gesture Recognition System for Sign Language. In *Proceedings of 3rd International Conference on Face an Gesture Recognition*. Nara, Japan, 558–567.

82. Pengfei Lu and Matt Huenerfauth. 2010. Collecting a motion-capture corpus of American Sign Language for data-driven generation research. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*. Association for Computational Linguistics, 89–97.

83. Silke Matthes, Thomas Hanke, Anja Regen, Jakob Storz, Satu Worseck, Eleni Efthimiou, Athanasia-Lida Dimou, Annelies Braffort, John Glauert, and Eva Safar. 2012. Dicta-Sign–building a multilingual sign language corpus. In *Proc. of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions Between Corpus and Lexicon (LREC), European Language Resources Association*. Istanbul.

84. Rachel I Mayberry and Robert Kluender. 2018. Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition* 21, 5 (2018), 886–905.

85. Carolyn McCaskill, Ceil Lucas, Robert Bayley, and Joseph Hill. 2011. The hidden treasure of Black ASL: Its history and structure. *Structure* 600 (2011), 83726.

86. Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

87. GRS Murthy and RS Jadon. 2009. A Review of Vision Vased Hand Gestures Recognition. *International Journal of Information Technology and Knowledge Management* 2, 2 (2009), 405–410.

88. University of Science and SLR Group Technology of China, Multimedia Computing & Communication. 2019. Chinese Sign Language Recognition Dataset. (2019). http://home.ustc.edu.cn/~pjh/dataset/cslr/index.html Accessed 2019-04-29.

89. World Federation of the Deaf. 2018. Our Work. (2018). http://wfdeaf.org/our-work/ Accessed 2019-03-26.

90. Mariusz Oszust and Marian Wysocki. 2013. Polish sign language words recognition with kinect. In *2013 6th International Conference on Human System Interactions (HSI)*. IEEE, 219–226.

91. Cemil Oz and Ming C. Leu. 2011. American Sign Language Word Recognition with a Sensory Glove Using Artificial Neural Networks. *Engineering Applications of Artificial Intelligence* 24, 7 (Oct. 2011), 1204–1213. DOI: http://dx.doi.org/10.1016/j.engappai.2011.06.015

92. Helen Petrie, Wendy Fisher, Kurt Weimann, and Gerhard Weber. 2004. Augmenting icons for deaf computer users. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1131–1134.

93. S. Prillwitz, R. Leven, H. Zienert, R. Zienert, and T. Hanke. 1989. *HamNoSys. Version 2.0*. Signum, Hamburg.

94. Jeanne Reis, Erin T Solovey, Jon Henner, Kathleen Johnson, and Robert Hoffmeister. 2015. ASL CLeaR: STEM education tools for deaf students. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. ACM, 441–442.

95. Jason Rodolitz, Evan Gambill, Brittany Willis, Christian Vogler, and Raja Kushalnagar. 2019. Accessibility of voice-activated agents for people who are deaf or hard of hearing. *Journal on Technology and Persons with Disabilities* 7 (2019).

96. Wendy Sandler. 2006. *Sign Language and Linguistic Universals*. Cambridge University Press.

97. Johan Schalkwyk, Doug Beeferman, Françoise Beaufays, Bill Byrne, Ciprian Chelba, Mike Cohen, Maryam Kamvar, and Brian Strope. 2010. Google search by voice: A case study. In *Advances in speech recognition*. Springer, 61–90.

98. Christoph Schmidt, Oscar Koller, Hermann Ney, Thomas Hoyoux, and Justus Piater. 2013a. Enhancing Gloss-Based Corpora with Facial Features Using Active Appearance Models. In *International Symposium on Sign Language Translation and Avatar Technology*, Vol. 2. Chicago, IL, USA.

99. Christoph Schmidt, Oscar Koller, Hermann Ney, Thomas Hoyoux, and Justus Piater. 2013b. Using Viseme Recognition to Improve a Sign Language Translation System. In *International Workshop on Spoken Language Translation*. Heidelberg, Germany, 197–203.

100. Jérémie Segouat and Annelies Braffort. 2009. Toward the study of sign language coarticulation: methodology proposal. In *Proceedings of the Second International Conferences on Advances in Computer-Human Interactions, Cancun, 2009.* 369–374. https://doi.org/10.1109/ACHI.2009.25

101. B. Shi, A. M. Del Rio, J. Keane, J. Michaux, D. Brentari, G. Shakhnarovich, and K. Livescu. 2018. American Sign Language Fingerspelling Recognition in the Wild. In *2018 IEEE Spoken Language Technology Workshop (SLT)*. 145–152. DOI: http://dx.doi.org/10.1109/SLT.2018.8639639

102. ShuR. 2013. SLinto Dictionary. (2013). http://slinto.com/us/ Accessed 2019-04-29.

103. LLC. Signing Savvy. 2019. SigningSavvy. (2019). https://www.signingsavvy.com/ Accessed 2019-05-02.

104. Robert Sparrow. 2005. Defending deaf culture: The case of cochlear implants. *Journal of Political Philosophy* 13, 2 (2005), 135–152.

105. T. Starner and A. Pentland. 1995. Real-Time American Sign Language Recognition from Video Using Hidden Markov Models. In *International Symposium on Computer Vision*. 265–270.

106. William C. Stokoe. 1960. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *Studies in linguistics: Occasional papers* 8 (1960).

107. W. C. Stokoe, D. Casterline, and C. Croneberg. 1965. *A Dictionary of American Sign Language on Linguistic Principles*. Linstok Press.

108. Jesus Suarez and Robin R Murphy. 2012. Hand Gesture Recognition with Depth Images: A Review. In *2012 IEEE RO-MAN: the 21st IEEE international symposium on robot and human interactive communication*. IEEE, 411–417.

109. V. Sutton and Deaf Action Committee for Sign Writing. 2000. *Sign Writing*. Deaf Action Committee (DAC).

110. Shinichi Tamura and Shingo Kawasaki. 1988. Recognition of Sign Language Motion Images. *Pattern Recognition* 21, 4 (1988), 343–353. DOI: http://dx.doi.org/10.1016/0031-3203(88)90048-9

111. Five Technologies. 2015. Five App. (2015). https://fiveapp.mobi/ Accessed 2019-04-29.

112. The Max Planck Institute for Psycholinguistics The language Archive. 2018. ELAN. (2018). https://tla.mpi.nl/tools/tla-tools/elan/elan-description/ Accessed 2019-04-29.

113. D. Uebersax, J. Gall, M. Van den Bergh, and L. Van Gool. 2011. Real-Time Sign Language Letter and Word Recognition from Depth Data. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 383–390. DOI: http://dx.doi.org/10.1109/ICCVW.2011.6130267

114. Els Van der Kooij. 2002. *Phonological Categories in Sign Language of the Netherlands: The Role of Phonetic Implementation and Iconicity*. Netherlands Graduate School of Linguistics.

115. Vcom3D. 2019. SigningAvatar. (2019). http://www.vcom3d.com/ Accessed 2019-04-29.

116. Tony Veale, Alan Conway, and Bróna Collins. 1998. The challenges of cross-modal translation: English-to-Sign-Language translation in the Zardoz system. *Machine Translation* 13, 1 (1998), 81–106.

117. Paranjape Ketki Vijay, Naphade Nilakshi Suhas, Chafekar Suparna Chandrashekhar, and Deshpande Ketaki Dhananjay. 2012. Recent developments in sign language recognition: A review. *Int. J. Adv. Comput. Eng. Commun. Technol* 1, 2 (2012), 21–26.

118. U. von Agris and K.-F. Kraiss. 2007. Towards a Video Corpus for Signer-Independent Continuous Sign Language Recognition. In *GW 2007 The 7th International Workshop on Gesture in Human-Computer Interaction and Simulation*, Sales Dias and Jota (Eds.). Lisbon, Portugal, 10–11.

119. Ronnie B Wilbur. 2000. Phonological and prosodic layering of nonmanuals in American Sign Language. *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (2000), 215–244.

120. Ying Wu and Thomas S Huang. 1999. Vision-based gesture recognition: A review. In *International Gesture Workshop*. Springer, 103–115.

121. Alexandros Yeratziotis. 2013. Sign Short Message Service (SSMS). (2013). http://www.ssmsapp.com/ Accessed 2019-04-29.

122. Zahoor Zafrulla, Helene Brashear, Peter Presti, Harley Hamilton, and Thad Starner. 2011a. CopyCat: an American sign language game for deaf children. In *Face and Gesture 2011*. IEEE, 647–647.

123. Zahoor Zafrulla, Helene Brashear, Thad Starner, Harley Hamilton, and Peter Presti. 2011b. American Sign Language Recognition with the Kinect. In *Proceedings of the 13th International Conference on Multimodal Interfaces (ICMI '11)*. ACM, New York, NY, USA, 279–286. DOI: http://dx.doi.org/10.1145/2070481.2070532

124. Morteza Zahedi, Philippe Dreuw, David Rybach, Thomas Deselaers, Jan Bungeroth, and Hermann Ney. 2006. Continuous Sign Language Recognition - Approaches from Speech Recognition and Available Data Resources. In *LREC Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios*. Genoa, Italy, 21–24.

125. Liwei Zhao, Karin Kipper, William Schuler, Christian Vogler, Norman Badler, and Martha Palmer. 2000. A machine translation system from English to American Sign Language. In *Conference of the Association for Machine Translation in the Americas*. Springer, 54–67.

Acronyms

AI Artificial Intelligence.

ASL American Sign Language.

CNN convolutional neural network.

DHH Deaf and hard of hearing.

HCI Human-Computer Interaction.

MT machine translation.

NLP Natural Language Processing.

UI/UX User Interface/User Experience.

WER Word Error Rate.

Glossary

annotation Descriptions of sign language content, typically in a writing system, note-taking system, or other discrete system that a computer can read.

ASL American Sign Language, the primary language of the Deaf community in the United States (and several other parts of the world).

audism Oppression or discrimination of the Deaf minority by the hearing majority. This marginalization has a long history, and may be overt or covert.

classifier A handshape used to represent a particular class of objects, actions, or ideas.

CNN A Convolutional Neural Network is a specific type of neural network that makes use of efficient parameter sharing and has revolutionized computer vision.

co-articulation The influence of surrounding signs/context on the production of the current sign, for example the ending of one sign affecting the beginning of the next.

continuous Refers to long phrases or full sentences as opposed to single, isolated signs (or words).

Deaf culture The culture of Deaf communities, of which sign languages are a central component. Capitalized "Deaf" refers to these cultures, while lowercase "deaf" refers to audiological status.

depiction Depiction refers to the ability to represent semantic information visually. Enacting an event or using classifiers are examples of depiction. Depiction is structured and highly flexible and many aspects of depiction have no spoken language corollary.

epenthesis Insertion of phonological features into or between signs, often in the form of movement for sign languages.

fingerspelling The use of a series of manual letters (handshapes representing letters), to spell out a written/spoken word in a sign language.

gloss Transliteration of one language (e.g., ASL) with words of another language (e.g., English). This retains the grammatical structure of the target language (e.g., ASL).

HamNoSys The Hamburg Notation System [93] is a phonetic transcription scheme that allows explicit annotation of individual subcomponents of sign language such as handshapes, movements, etc.

phonological Having to do with phonology, the smallest units of a language, and how they are put together. For spoken languages the components are sounds, and for sign languages they are handshapes, locations, and movements etc.

real-life Describes sign language that has not been scripted or tailored to suit machine learning algorithms, but is rather produced naturally.

signer-independent Characterizes a realistic sign language recognition experimental setup where the signers that are tested on are explicitly excluded from the training set.

User Interface/User Experience UI refers to user interface, while UX refers to user experience. Technology designers (and HCI researchers) are often trained to focus on these crucial aspects of technology.

vocabulary The set of unique signs (or words) that occur in a language or dataset.

Word Error Rate A standard evaluation criterion in speech recognition and sign language recognition, and is based on the edit distance between reference and hypothesis.