Universal Access in the Information Society
https://doi.org/10.1007/s10209-023-00992-1

REVIEW PAPER

Machine translation from signed to spoken languages: state of the art and challenges

Mathieu De Coster[1] · **Dimitar Shterionov**[2] · **Mieke Van Herreweghe**[3] · **Joni Dambre**[1]

Accepted: 21 March 2023

Abstract
Automatic translation from signed to spoken languages is an interdisciplinary research domain on the intersection of computer vision, machine translation (MT), and linguistics. While the domain is growing in terms of popularity—the majority of scientific papers on sign language (SL) translation have been published in the past five years—research in this domain is performed mostly by computer scientists in isolation. This article presents an extensive and cross-domain overview of the work on SL translation. We first give a high level introduction to SL linguistics and MT to illustrate the requirements of automatic SL translation. Then, we present a systematic literature review of the state of the art in the domain. Finally, we outline important challenges for future research. We find that significant advances have been made on the shoulders of spoken language MT research. However, current approaches often lack linguistic motivation or are not adapted to the different characteristics of SLs. We explore challenges related to the representation of SL data, the collection of datasets and the evaluation of SL translation models. We advocate for interdisciplinary research and for grounding future research in linguistic analysis of SLs. Furthermore, the inclusion of deaf and hearing end users of SL translation applications in use case identification, data collection, and evaluation, is of utmost importance in the creation of useful SL translation models.

Keywords Sign language · Computer vision · Machine translation · Deep learning · Literature review

1 Introduction

The speedy progress in deep learning has seemingly enabled a bevy of new applications related to sign language recognition, translation, and synthesis, which can be grouped under the umbrella term "sign language processing." Sign Language Recognition (SLR) can be likened to "information extraction from sign language data," for example fingerspelling recognition [1, 2] and sign classification [3, 4]. Sign Language Translation (SLT) maps this extracted information to meaning and translates it to another (signed or spoken) language [5, 6]; the opposite direction, from text to sign language, is also possible [7, 8]. Sign Language Synthesis (SLS) aims to generate sign language from some representation of meaning, for example through virtual avatars [9, 10]. In this article, we are zooming in on translation from signed languages to spoken languages.

In particular, we focus on translating videos containing sign language utterances to text, i.e., the written form of spoken language. We will only discuss SLT models that support video data as input, as opposed to models that require wearable bracelets or gloves, or 3D cameras. Systems that use smart gloves, wristbands or other wearables are considered intrusive and not accepted by sign language communities (SLCs) [11]. In addition, they are unable to capture all information present in signing, such as non-manual actions. Video-based approaches also have benefits compared to wearable-based approaches: they can be trained with existing data, and they could for example be integrated into

✉ Mathieu De Coster
mathieu.decoster@ugent.be

Dimitar Shterionov
d.shterionov@tilburguniversity.edu

Mieke Van Herreweghe
mieke.vanherreweghe@ugent.be

Joni Dambre
joni.dambre@ugent.be

[1] IDLab-AIRO, Ghent University - imec, Technologiepark-Zwijnaarde 126, Ghent 9052, Belgium

[2] Tilburg University, Warandelaan 2, Tilburg 5037 AB, Netherlands

[3] Ghent University, Blandijnberg 2, Ghent 9000, Belgium

Published online: 01 April 2023

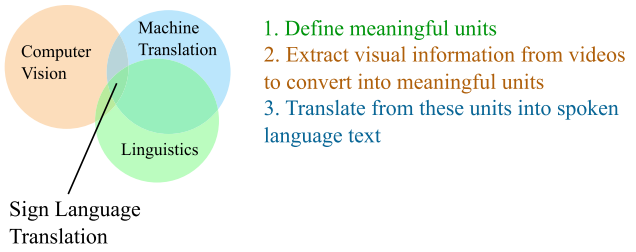


Fig. 1 Sign language translation lies on the intersection of computer vision, machine translation, and linguistics

conference calling software, or used for automatic captioning in videos of signing vloggers.

Several previously published scientific papers liken SLR to gesture recognition (among others, [12, 13]), or even present a fingerspelling recognition system as an SLT solution (among others, [14, 15]). Such classifications are overly simplified and incorrect. They may lead to a misunderstanding of the technical challenges that must be solved. As Fig. 1 illustrates, SLT lies on the intersection of computer vision, machine translation, and linguistics. Experts from each domain must come together to truly address SLT.

This article aims to provide a comprehensive overview of the state of the art (SOTA) of sign to spoken language translation. To do this, we perform a systematic literature review and discuss the state of the domain. We aim to find answers to the following research questions:

RQ1. Which datasets are used, for what languages, and what are the properties of these datasets?

RQ2. How should we represent sign language data for Machine Translation (MT) purposes?

RQ3. Which algorithms are currently the SOTA for SLT?

RQ4. How are current SLT models evaluated?

Furthermore, we list several challenges in SLT. These challenges are of a technical and linguistic nature. We propose research directions to tackle them.

In parallel with this article, another survey on SLT was written and published [16]. It provides a narrative historical overview of the domains of SLR and SLT and positions them in the wider scope of sign language processing. They also discuss the "to-sign" direction of SLT that we disregard. We provide a systematic and extensive analysis of the most recent work on SLT, supported by a discussion of sign language linguistics. Their work is a broader overview of the domain, but less in depth and remains mostly limited to computer science.

This article is also related to the work of Bragg et al. [17], that gives a limited but informative overview of the domains of SLR, SLS and SLT that is based on the results of panel discussions. They list several challenges in the field that align with our own findings, e.g., data scarcity and involvement of SLCs.

We first provide a high level overview of some required background information on sign languages in Sect. 2. This background can help in the understanding of the remainder of this article, in particular the reasoning behind our inclusion criteria. Section 3 provides the necessary background in machine translation. We discuss the inclusion criteria and search strategy for our systematic literature search in Sect. 4 and objectively compare the considered papers on SLT in Sect. 5; this includes Sect. 5.7 focusing on a specific benchmark dataset. The research questions introduced above are answered in our discussion of the literature overview, in Sect. 6. We present several open challenges in SLT in Sect. 7. The conclusion and takeaway messages are given in Sect. 8.

2 Sign language background

2.1 Introduction

It is a common misconception that there exists a single, universal, sign language. Just like spoken languages, sign languages evolve naturally through time and space. Several countries have national sign languages, but often there are also regional differences and local dialects. Furthermore, signs in a sign language do not have a one-to-one mapping to words in any spoken language: translation is not as simple as recognizing individual signs and replacing them with the corresponding words in a spoken language. Sign languages have distinct vocabularies and grammars and they are not tied to any spoken language. Even in two regions with a shared spoken language, the regional sign languages used can differ greatly. In the Netherlands and in Flanders (Belgium), for example, the majority spoken language is Dutch. However, Flemish Sign Language (VGT) and the Sign Language of the Netherlands (NGT) are quite different. Meanwhile, VGT is linguistically and historically much closer to French Belgian Sign Language (LSFB) [18], the sign language used primarily in the French-speaking part of Belgium, because both originate from a common Belgian Sign Language, diverging in the 1990s [19]. In a similar vein, American Sign Language (ASL) and British Sign Language (BSL) are completely different even though the two countries share English as the official spoken language.

2.2 Sign language characteristics

2.2.1 Sign components

Sign languages are visual; they make use of a large space around the signer. Signs are not composed solely of manual gestures. In fact, there are many more components to a sign. Stokoe stated in 1960 that signs are composed of hand shape, movement and place of articulation parameters [20]. Battison later added orientation, both of the palm and of the fingers [21]. There are also non-manual components such as mouth patterns. These can be divided into mouthings—where the pattern refers to (part of) a spoken language word—and mouth gestures, e.g., touting one's lips. Non-manual components play an important role in sign language lexicons and grammars [22]. They can, for example, separate minimal pairs: signs that share all articulation parameters but one. When hand shape, orientation, movement and place of articulation are identical, mouth patterns can for example be used to differentiate two signs. Non-manual actions are not only important at the lexical level as just illustrated, but also at the grammatical level. A clear example of this can be found in eyebrow movements: furrowing or raising the eyebrows can signal that a question is being asked and indicate the type of question (open or closed).

2.2.2 Simultaneity

Sign languages exhibit simultaneity on several levels. There is simultaneity on the component level: as explained above, manual actions can be combined with non-manual actions simultaneously. We also observe simultaneity at the utterance level. It is, for example, possible to turn a positive utterance into a negative utterance by shaking one's head while performing the manual actions. Another example is the use of eyebrow movements to transform a statement into a question.

2.2.3 Signing space

The space around the signer can also be used to indicate, for instance, the location or moment in time of the conversational topic. A signer can point behind their back to specify that an event occurred in the past and likewise, point in front of them to indicate a future event. An imaginary timeline can also be constructed in front of the signer, with time passing from left to right. Space is also used to position referents [18, 23]. For example, a person can be discussing a conversation with their mother and father. Both referents get assigned a location (locus) in the signing space and further references to these persons are made by pointing to, looking at, or signing toward these loci. For example, "mom gives something to dad" can be signed by moving the sign for "to give" from the locus associated with the mother to the one associated with the father. Modeling space, detecting positions in space, and remembering these positions is important for SLT models.

2.2.4 Classifiers

Another important aspect of sign languages is the use of classifiers. Zwitserlood [24] describes them as "morphemes with a non-specific meaning, which are expressed by particular configurations of the manual articulator (or: hands) and which represent entities by denoting salient characteristics." There are many more intricacies of classifiers than can be listed here, so we give a limited set of examples instead. Several types of classifiers exist. They can, for example, represent nouns or adjectives according to their shape or size. Whole entity classifiers can be used to represent objects, e.g., a flat hand can represent a car; handling classifiers can be used to indicate that an object is being handled, e.g., a pencil is picked up from a table. In a whole entity classifier, the articulator represents the object, whereas in a handling classifier it operates on the object.

2.2.5 The established and the productive lexicon

The vocabularies of sign languages are not fixed. Oftentimes new signs are constructed by sign language users. On the one hand, sign languages can borrow signs from other sign languages, similar to loanwords in spoken languages. In this case, these signs become part of the established lexicon. On the other hand, there is the productive lexicon—one can create an ad hoc sign. Vermeerbergen [25] gives the example of "a man walking on long legs" in VGT: rather than expressing this clause by signing "man," "walk," "long" and "legs", the hands are used (as classifiers) to imitate the man walking. Both the established and productive lexicons are integral parts of sign languages.

Signers can also enact other subjects with their whole body, or part of it. They can, for example, enact animals by imitating their movements or behaviors.

2.2.6 Fingerspelling

Fingerspelling can be used to convey concepts for which a sign does not (yet) exist, or to introduce a person who has not yet been assigned a name sign. It is based on the alphabet of a spoken language, where every letter in that alphabet has a corresponding (static or dynamic) sign. Fingerspelling is also not shared between sign languages. For example, in ASL, fingerspelling is one-handed, but in BSL two hands are used.

2.3 Notation systems for sign languages

Unlike many spoken languages, sign languages do not have a standardized written form. Several notation systems do exist, but none of them are generally accepted as a standard [26]. The earliest notation system was proposed in the 1960s, namely the Stokoe notation [20]. It was designed for ASL and comprises a set of symbols to notate the different components of signs. The position, movement and orientation of the hands are encoded in iconic symbols, and for hand shapes, letters from the Latin alphabet corresponding to the most similar fingerspelling hand shape are used [20]. Later, in the 1970s, Sutton introduced SignWriting[1]: a notation system for sign languages based on a dance choreography notation system [27]. The SignWriting notation for a sign is composed of iconic symbols for the hands, face and body. The signing location and movements are also encoded in symbols, in order to capture the dynamic nature of signing. SignWriting is designed as a system for writing signed utterances for everyday communication. In 1989, the Hamburg Notation System (HamNoSys) was introduced [28]. Unlike SignWriting, it is designed mainly for linguistic analysis of sign languages. It encodes hand shapes, hand orientation, movements and non-manual components in the form of symbols.

Stokoe notation, SignWriting and HamNoSys represent the visual nature of signs in a compact format. They are notation systems that operate on the phonological level. These systems, however, do not capture the meaning of signs. In linguistic analysis of sign languages, glosses are typically used to represent meaning. A sign language gloss is a written representation of a sign in one or more words of a spoken language, commonly the majority language of the region. Glosses can be composed of single words in the spoken language, but also of combinations of words. Examples of glosses are: "CAR," "BRIDGE," but also "car-crosses-bridge." Glosses do not accurately represent the meaning of signs in all cases and glossing has several limitations and problems [26]. They are inherently sequential, whereas signs often exhibit simultaneity [29].[2] Furthermore, as glosses are based on spoken languages, there may be an implicit influence of the spoken language projected onto the sign language [25, 26]. Finally, there is no universal standard on how glosses should be constructed: this leads to differences between corpora of different sign languages, or even between several sign language annotators working on the same corpus [30].

Sign_A is a recently developed framework that aims to define an architecture that is sufficiently robust to model sign languages on both the phonological level as well as containing meaning (when combined with a role and reference grammar (RRG)) [31]. Sign_A with RRG does not only encode the meaning of sign language utterances, but also parameters pertaining to manual and non-manual actions. De Sisto et al. [32] propose investigating the application of Sign_A for data-driven SLT systems.

The above notation systems for sign languages range from graphical to written and computational representations of signs and signed utterances. None of these notation systems were originally designed for the purpose of automatic translation from signed to spoken languages, but they can be used to train MT models. For example, glosses are often used for SLT because of their similarity to written language text, e.g., [5, 6]. These notation systems can also be used as labels to pre-train feature extractors for SLT models. For instance, Koller et al. presented SLR systems that exploit SignWriting [33, 34], and these systems are leveraged in some later works on SLT, e.g., [35, 36]. Many SLT models also use feature extractors that were pre-trained with gloss labels, e.g., [37, 38].

3 Machine translation

3.1 Spoken language MT

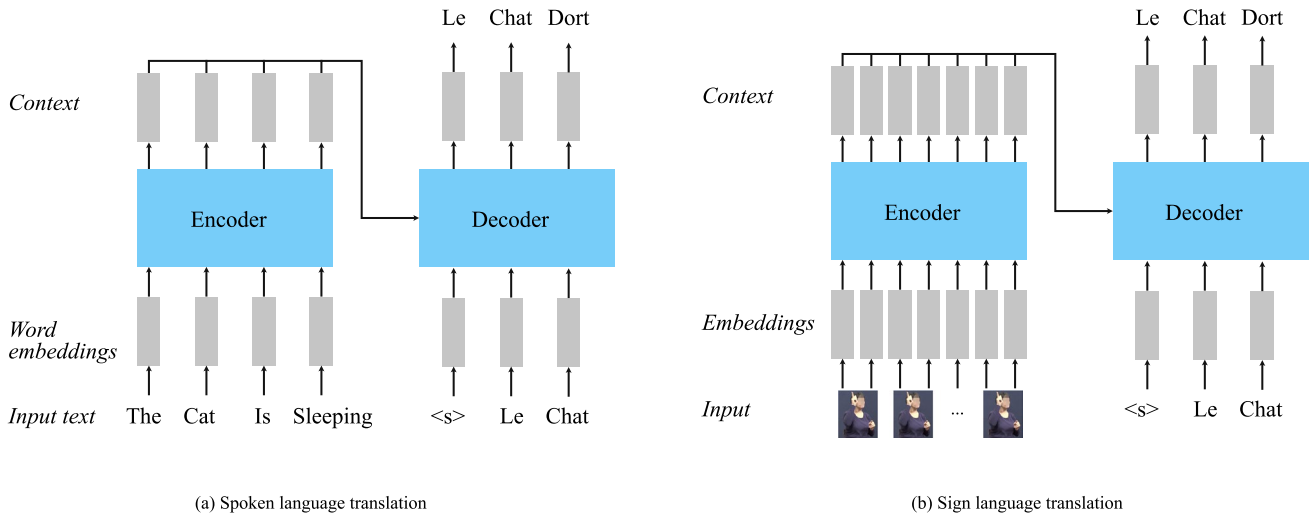
Machine translation is a sequence-to-sequence task. That is, given an input sequence of tokens that constitute a sentence in a source language, an MT system generates a new sequence of tokens that represents a sentence in a target language. A token refers to a sentence construction unit: a word, a number, a symbol, a character or a subword unit.

Current SOTA models for spoken language MT are based on a neural encoder-decoder architecture: an encoder network encodes an input sequence in the source language into a multi-dimensional representation; it is then fed into a decoder network which generates a hypothesis translation conditioned on this representation. The original encoder-decoder was based on Recurrent Neural Networks (RNNs) [39]. To deal with long sequences, Long Short-Term Memory Networks (LSTMs) [40] and Gated Recurrent Units (GRUs) [41] were used. To further improve the performance of RNN-based MT, an attention mechanism was introduced by Bahdanau et al. [42]. In recent years the transformer architecture [43], based primarily on the idea of attention (in combination with positional encoding) has pushed the SOTA even further.

As noted above, a sentence is broken down into tokens and each token is fed into the Neural Machine Translation (NMT) model. NMT converts each token into a

[1] https://signwriting.org/.

[2] For this reason, annotators of sign language corpora sometimes provide two parallel gloss tiers: one per hand [30].



(a) Spoken language translation

(b) Sign language translation

Fig. 2 Neural machine translation models for spoken (**a**) and sign (**b**) language translation are similar; the main difference is the input modality: text for (**a**) and video for (**b**)

multidimensional representation before that token representation is used in the encoder or decoder to construct a sentence level representation. These token representations, typically referred to as word embeddings, encode the meaning of a token based on its context. Learning word embeddings is a monolingual task, since they are associated with tokens in a particular language. Given that for a large number of languages and use cases monolingual data is abundant, it is relatively easy to build word embedding models of high quality and coverage. Building such word embedding models is typically performed using unsupervised algorithms such as GLoVe [44], BERT [45] and BART [46]. These algorithms encode words into vectors in such a way that the vectors of related words are similar.[3]

The domain of spoken language MT is extensive and the current SOTA of NMT builds upon years of research. To provide a complete overview of spoken language MT is out of scope for this article. For a more in depth overview of the domain, we refer readers to the work of Stahlberg [49].

3.2 Sign language MT

Conceptually, sign language MT and spoken language MT are similar. The main difference is the input modality. Spoken language MT operates on two streams of discrete tokens (text to text). As sign languages have no standardized notation system, a generic SLT model must translate from a continuous stream to a discrete stream (video to text). To reduce the complexity of this problem, sign language videos are discretized to a sequence of still frames that make up the video. SLT can now be framed as a sequence-to-sequence, frame-to-token task. As they are, these individual frames do not convey meaning in the way that the word embeddings in a spoken language translation model do. Even though it is possible to train SLT models using frame-based representations as inputs, the extraction of salient sign language representations is required to facilitate the modeling of meaning in sign language encoders.

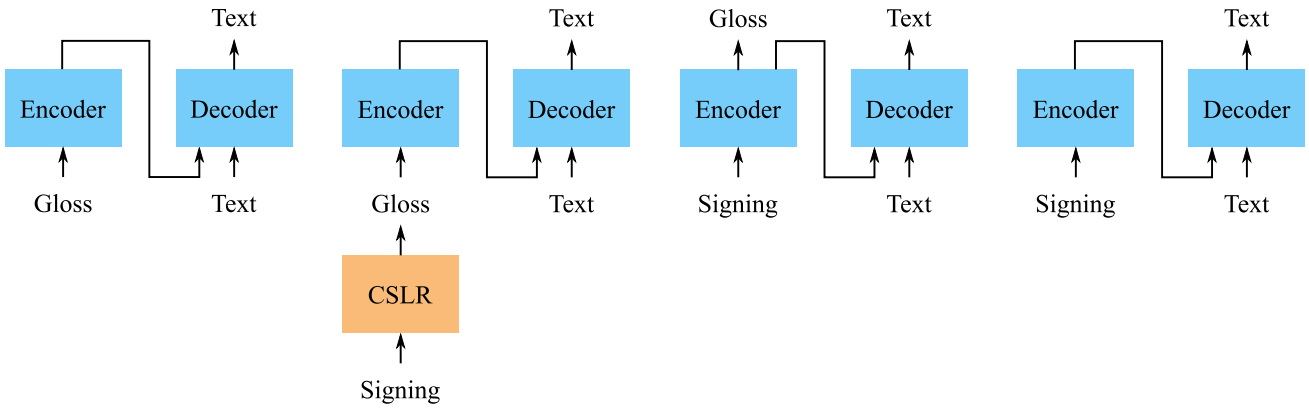
Figure 2 shows a spoken language NMT and sign language NMT model side by side. The main difference between the two is the input modality. For a spoken language NMT model, both the inputs and outputs are text. For a sign language NMT model, the inputs are some representation of sign language (in this case, video embeddings). Other than this input modality, the models function similarly and are trained and evaluated in the same manner.

3.2.1 Sign language representations

For the encoder of the translation model to capture the meaning of the sign language utterance, a salient representation for sign language videos is required. We can differentiate between representations that are linked to the source modality, namely videos, and linguistically motivated representations.

As will be discussed in Sect. 5.4, the former type of representations are often frame-based, i.e., every frame in the video is assigned a vector, or clip-based, i.e., clips of arbitrary length are assigned a vector. These types of representations are rather simple to derive, e.g., by extracting

[3] According to the Distributional Semantics, words that have the same or similar meaning appear in the same context and as such the meaning of a word can be defined by the context in which it appears [47, 48].



(a) Gloss2Text (b1) Sign2Gloss2Text (b2) Sign2Gloss, Gloss2Text (c) Sign2(Gloss+Text) (d) Sign2Text

Fig. 3 There are five distinct translation tasks in the considered scientific literature on SLT. The Sign2Gloss2Text and (Sign2Gloss, Gloss2Text) models both use the same architecture, but a different training algorithm

information directly from a Convolutional Neural Network (CNN). However, they suffer from two main drawbacks. First, such representations are fairly long. For example, the RWTH-PHOENIX-Weather 2014T dataset [6] contains samples of on average 114 frames (in German Sign Language (DGS)), whereas the average sentence length (in German) is 13.7 words in that dataset. As a result, frame-based representations for sign languages negatively impact the computational performance of SLT models. Second, such representations do not originate from domain knowledge. That is, they do not capture the semantics of sign language. If semantic information is not encoded in the sign language representation, the translation model is forced to model the semantics and perform translation at the same time.

The second category includes a range of linguistically motivated representations, from semantic representations to individual sign representations. In Sect. 2.3, we presented an overview of some notation systems for sign languages: Stokoe notation, SignWriting, HamNoSys, glosses, and Sign_A. These notation systems can be used as representations in an SLT model, or to pre-train the feature extractor of SLT models. In current research, only glosses have been used as inputs or labels for the SLT models themselves, because large annotated datasets for the other systems do not exist.

3.2.2 Tasks

The reviewed papers cover five distinct translation tasks that can be classified based on whether, and how, glosses are used. To denote these tasks, we borrow the naming conventions from Camgöz et al. [6, 37]. These tasks are illustrated in Fig. 3.

Gloss2Text Gloss2Text models are used to translate from sign language glosses to spoken language text. They provide a reference for the performance that can be achieved using a salient representation. Therefore they can serve as a compass for the design of sign language representations and the corresponding SLR systems. Note that the performance of a Gloss2Text model is not an upper bound for the performance of an SLT model: glosses do not capture all linguistic properties of signs (see Sect. 2.3).

Sign2Gloss2Text A Sign2Gloss2Text translation system includes an SLR system as the first step, to predict glosses from video. Consequently, errors made by the recognition system are propagated to the translation system. Camgöz et al. [6] for example report a drop in translation accuracy when comparing a Sign2Gloss2Text system to a Gloss2Text system.

(Sign2Gloss, Gloss2Text) In this training setup, first a Gloss2Text model is trained using ground truth gloss and text data. Then, this model is fixed and used to evaluate the performance of the entire translation model, including the SLR model (Sign2Gloss). This is different from Sign2Gloss-2Text models, where the Gloss2Text model is trained with the gloss annotations generated by the Sign2Gloss model. Camgöz et al. [6] show that these models perform worse than Sign2Gloss2Text models, because those can learn to correct the noisy outputs of the Sign2Gloss model in the translation model.

Sign2(Gloss+Text) Glosses can provide a supervised signal to a translation system without being an information bottleneck, if the model is trained to jointly predict both glosses and text [37]. Such a model must be able to predict glosses and text from a single sign language representation. The gloss labels provide additional information to the encoder, facilitating the training process. In a Sign2Gloss2Text model

(as previously discussed), the translation model receives glosses as inputs: any information that is not present in glosses cannot be used to translate into spoken language text. In Sign2(Gloss+Text) models, however, the translation model input is the sign language representation (embeddings), which may be richer.

Sign2Text Sign2Text models forgo the explicit use of a separate SLR model, and instead perform translation directly with features that are extracted from videos. Sign2Text models do not need glosses to train the translation model. Note that in some cases, these features are still extracted using a model that was pre-trained for SLR, e.g., [37, 38]. This means that some Sign2Text models do indirectly require gloss level annotations for training.

3.3 Requirements for sign language MT

With the given information on sign language linguistics and MT techniques, we are now able to sketch the requirements for sign language MT.

3.3.1 Data requirements

The training of data-driven MT models requires large datasets. The collection of such datasets is expensive and should therefore be tailored to specific use cases. To determine these use cases, members of SLCs must be involved. We answer RQ1 by providing an overview of existing datasets for SLT.

3.3.2 Video processing and sign language representation

We need to be able to process sign language videos and convert them into an internal representation (SLR). This representation must be rich enough to cover several aspects of sign languages (including manual and non-manual actions, simultaneity, signing space, classifiers, the productive lexicon, and fingerspelling). We look in our literature overview for an answer to RQ2 on how we should represent sign language data.

3.3.3 Translating between sign and spoken representations

We need to be able to translate from such a representation into a spoken language representation, which can be reused from existing spoken language MT systems. We need to adapt NMT systems to be able to work with the sign language representation, which will possibly contain simultaneous elements. By comparing different methods for SLT, we evaluate which MT algorithms perform best in the current SOTA (RQ3).

3.3.4 Evaluation

The evaluation of the resulting models can be automated by computing metrics on corpora. These metrics provide an estimate of the quality of translations. Human evaluation (by hearing and deaf people, signing and non-signing) and qualitative evaluations can provide insights into the models and data. We illustrate how current SLT models are evaluated (RQ4).

4 Literature review methodology

4.1 Inclusion criteria and search strategy

To provide an overview of sound SLT research, we adhere to the following principles in our literature search. We consider only peer-reviewed publications. We include journal articles as well as conference papers: the latter are especially important in computer science research. Any paper that is included must be on the topic of sign language machine translation and must not misrepresent the natural language status of sign languages. Therefore, we omit any papers that present classification or transcription of signs or fingerspelling recognition as SLT models (we will show in this section that there are many papers that do this). As we focus on non-intrusive translation from sign languages to text, we exclude papers that use gloves or other wearable devices.

Three scientific databases were queried: Google Scholar, Web of Science and IEEE Xplore.[4] Four queries were used to obtain initial results: "sign language translation," "sign language machine translation," "gloss translation" and "gloss machine translation." These key phrases were chosen for the following reasons. We aimed to obtain scientific research papers on the topic of MT from sign to spoken languages; therefore, we search for "sign language machine translation." Several works perform translation between sign language glosses and spoken language text, hence "gloss machine translation". As many papers omit the word "machine" in "machine translation," we also include the key phrases "sign language translation" and "gloss translation."

4.2 Search results and selection of papers

Our initial search yielded 855 results, corresponding to 565 unique documents. We applied our inclusion criteria step by step (see Table 1), and obtained a final set of 57 papers [5, 6,

[4] Google Scholar: https://scholar.google.com, Web of Science: https://www.webofscience.com, IEEE Xplore: https://ieeexplore.ieee.org/.

Table 1 Application of the inclusion criteria to the initial search results

Step	Criterion	Excluded	Remaining
1. All search results	Match search queries	–	855
2. Unique search results	No duplicate results	290	565
3. English documents	English	29	536
4. Peer-reviewed papers	Peer-reviewed, paper	101	435
5. Sign language translation papers	Related to sign language	30	405
	On sign language processing	60	345
	No fingerspelling	52	293
	No sign classification	58	235
	No sign language recognition	20	215
	Implements machine translation	36	179
6. Sign language to spoken language	Sign to spoken translation	117	62
7. Video-based	No gloves or other wearables	5	57

35–38, 50–100]. The complete list of search results can be found in supplementary material (resource 1).

We further explain the reasons for excluding papers with examples. We found 30 papers not related to sign language. These papers discuss the classification and translation of traffic signs and other public signs. 60 papers consider a topic related to sign language, but not to sign language processing. These include papers from the domains of linguistics and psychology. Out of the remaining 345 papers, 130 papers claim to present a translation model in their title or main text, but in fact present a fingerspelling recognition (52 papers [14, 15, 101–150]), sign classification (58 papers [151–208]), or SLR (20 papers [209–228]) system. There are 36 papers ([16, 30, 229–262]) on various topics within the domain of sign language processing that do not implement a new MT model.

We find double the amount of papers on MT from spoken languages to sign languages than vice versa: 117 compared to 59. These papers are closely related to the subject of this article, but often use different techniques, including virtual avatars (e.g., [9, 10]), due to the different source and target modality. Hence, translation from a spoken language to a sign language is outside the scope of this article. For an overview of this related field, we refer readers to a recent review article by Kahlon and Singh [263].

Remark that our final inclusion criterion, "present a non-intrusive system based only on RGB camera inputs" is almost entirely covered by the previous criteria. We find several papers that present glove-based systems, but they do not present translation systems. Instead, they are focused on fingerspelling recognition or sign classification (e.g., [119, 164–166]). The following five papers present an intrusive SLT system. Fang et al. [264] present an SLT system where signing deaf and hard of hearing people wear a device with integrated depth camera and augmented reality glasses to communicate with hearing people. Guo et al. [265] use a Kinect (RGB-D) camera to record the sign language data. The data used by Xu et al. [266] was also recorded using a Kinect. Gu et al. propose wearable sensors [267] and so do Zhang et al. [268]. After discarding these five papers, we obtain the final set of 57.

5 Literature overview

5.1 Sign language MT

Following our methodology on paper selection, laid out in Sect. 4, we obtain 57 papers published from 2004 until and including 2022. In the analysis, papers are classified based on tasks, datasets, methods and evaluation techniques.

The early work on MT from signed to spoken languages is based entirely on statistical methods [5, 50–58]. These works focus on gloss based translation. Several of them add visual inputs to augment the (limited) information provided by the glosses. Bungeroth et al. present the first statistical model that translates from signed to spoken languages [5]. They remark that glosses have limitations and need to be adapted for use in MT systems. Stein et al. incorporate visual information in the form of small images and hand tracking information to augment their model and enhance its performance [50], as do Dreuw et al. [51]. Dreuw et al. later ground this approach by listing requirements for SLT models, such as modeling simultaneity, signing space, and handling coarticulation [53]. Schmidt et al. further add non-manual visual information by incorporating lip reading [57]. The other papers in this set use similar techniques but on different datasets, or compare SMT algorithms [52, 54–56, 58].

In 2018, the domain moved away from SMT and toward NMT. This trend is clearly visible in Fig. 4. This drastic shift was not only motivated by the successful applications of NMT techniques in spoken language MT, but also by the publication of the RWTH-PHOENIX-Weather 2014T dataset and the promising results obtained on that dataset using

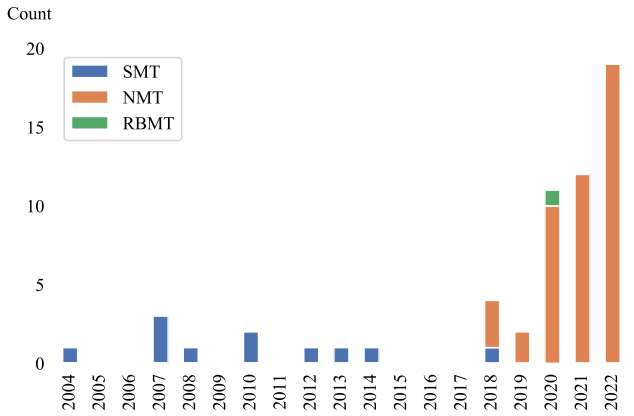


Fig. 4 The earlier papers on SLT all propose Statistical Machine Translation (SMT) models, but since 2018, NMT has become the dominant variant

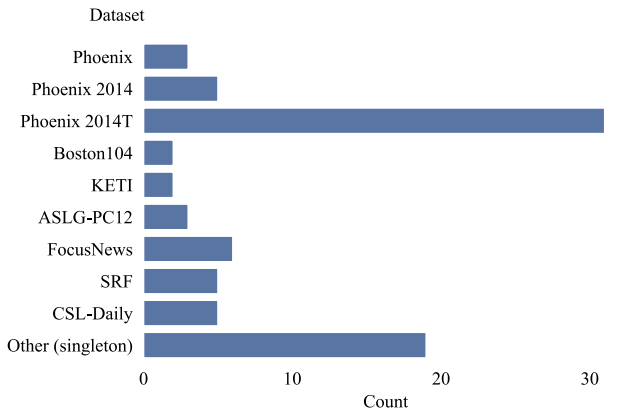


Fig. 5 The RWTH-PHOENIX-Weather 2014T dataset is used the most (31 times) throughout literature whereas other datasets are referenced at most six times in the 57 discussed papers

NMT methods [6]. Two exceptions are found. Luqman et al. [63] use Rule-based Machine Translation (RBMT) in 2020 to translate from Arabic sign language into Arabic, and Moe et al. [60] compare NMT and SMT approaches for Myanmar sign language translation in 2018.

Between 2004 and 2018, research into translation from signed to spoken languages was sporadic (10 papers in our subset were published over 14 years). Since 2018, with the move toward NMT, the domain has become more popular, with 47 papers in our subset published over the span of 5 years.

5.2 Datasets

Several datasets are used in SLT research. Some are used often, whereas others are only used once. The distribution is shown in Fig. 5. It is clear that the most used dataset is RWTH-PHOENIX-Weather 2014T [6]. This is because it was the first dataset large enough for neural SLT and because it is readily available for research purposes. This dataset is an extension of earlier versions, RWTH-PHOENIX-Weather [269] and RWTH-PHOENIX-Weather 2014 [58]. It contains videos in DGS, gloss annotations, and text in German. Precisely because of the popularity of this dataset, we can compare several approaches to SLT: see Sect. 5.7.

Other datasets are also used several times. The KETI dataset [62] contains Korean Sign Language (KSL) videos, gloss annotations, and Korean text. RWTH-Boston-104 [50] is a dataset for ASL to English translation containing ASL videos, gloss annotations, and English text. The ASLG-PC12 dataset [270] contains ASL glosses and English text. The glosses are generated from English with a rule based approach. FocusNews and SRF were both introduced as part of the WMTSLT22 task [85], and they contain news broadcasts in Swiss German Sign Language (DSGS) with German translations. CSL-Daily is a dataset containing translations from Chinese Sign Language (CSL) to Chinese on everyday topics [77].

In 2022, the first SLT dataset containing parallel data in multiple sign languages was introduced [98]. The SP-10 dataset contains sign language data and parallel translations from ten sign languages. It was created from data collected in the SpreadTheSign research project [271]. Yin et al. [98] show that multilingual training of SLT models can improve performance and allow for zero-shot translation.

Several papers use the CSL dataset to evaluate SLT models [72, 79, 94]. However, this is problematic because this dataset was originally proposed for SLR [272]. Because the sign (and therefore gloss) order is the same as the word order in the target spoken language, this dataset is not suited for the evaluation of translation models (as explained in Sect. 2.1).

Table 2 presents an overview of dataset and vocabulary sizes. The number of sign instances refers to the amount of individual signs that are produced. Each of these belongs to a vocabulary, of which the size is also given. Finally, there can be singleton signs: these are signs that occur only in the training set but not in the validation or test sets. ASLG-PC12 contains 827 thousand training sentences. It is the largest dataset in terms of number of parallel sentences. The most popular dataset with video data (RWTH-PHOENIX-Weather 2014T) contains only 7,096 training sentences. For MT between spoken languages, datasets typically contain several millions of sentences, for example the Paracrawl corpus [273]. It is clear that compared to spoken language datasets, sign language datasets lack labeled data. In other words, SLT is a low-resource MT task.

Table 2 Statistics of the datasets that are used in more than one paper

Dataset	Languages	Sentences	Sign instances	Vocab. size	Singletons
Phoenix [269]	DGS-German	3118	25,449	768	248
Phoenix 2014 [58]	DGS-German	9015	102,726	1580	565
Phoenix 2014T [6]	DGS-German	8257	75,783	1870	337
Boston-104 [50]	ASL-English	201	888	168	27
KETI [62]	KSL-Korean	14,672	–	524	–
ASLG-PC12 [270]	ASL-English	87,709	913,579	22,255	6133
CSL-Daily [77]	CSL-Chinese	20,654	–	2000	–
FocusNews [85]	DSGS-German	10,136	–	–	–
SRF [85]	DSGS-German	7071	–	–	–

5.3 Tasks

A total of 20 papers report on a Gloss2Text model [5, 6, 37, 51, 52, 54–58, 60, 61, 63, 65, 68, 70, 71, 74, 91, 100]. Sign-2Gloss2Text models are proposed in five papers [6, 37, 65, 77, 94] and (Sign2Gloss, Gloss2Text) models also in five [6, 37, 50, 51, 65]. Sign2(Gloss+Text) models are found eight times within the reviewed papers [37, 38, 78, 80, 88, 89, 92, 99] and Sign2Text models 28 times [6, 35–37, 62, 64, 66, 67, 69, 72, 73, 75, 76, 79, 81–87, 90, 92, 93, 95–98].

Before 2018, when SMT was dominant, Gloss2Text models were most popular, being proposed nine times out of eleven models, the other two being (Sign2Gloss, Gloss-2Text) models. Since 2018, with the availability of larger datasets, deep neural feature extractors and neural SLR models, Sign2Gloss2Text, Sign2(Gloss+Text) and Sign2Text are becoming dominant. This gradual evolution from Gloss-2Text models toward end-to-end models is visible in Fig. 6.[5]

5.4 Sign language representations

The sign language representations used in the current scientific literature are glosses and representations extracted from videos. Early on, researchers using SMT models for SLT recognized the limitations of glosses and began to add additional visual information to their models [50, 51, 53, 57]. The advent of CNNs has made processing and incorporating visual inputs easier and more robust. All but one model since 2018 that include feature extraction, use neural networks to do so.

We examine the representations on two dimensions. First, there is the method of extracting visual information (e.g., by using human pose estimation or CNNs). Second, there is the matter of *which* visual information is extracted (e.g., full frames, or specific parts such as the hands or face).

5.4.1 Extraction methods

The most popular feature extraction method in modern SLT is the 2D CNN. 20 papers use a 2D CNN as feature extractor [6, 35–38, 64, 65, 72, 75, 77–81, 87, 88, 92, 93, 95, 98]. These are often pre-trained for image classification using the ImageNet dataset [274]; some are further pre-trained on the task of Continuous Sign Language Recognition (CSLR), e.g., [37, 38, 92, 93]. Three papers use a subsequent 1D CNN to temporally process the resulting spatial features [64, 77, 80].

Human pose estimation systems are used to extract features in fifteen papers [35, 36, 62, 69, 73, 79, 80, 84, 85, 88–90, 94, 95, 97]. The estimated poses can be the only inputs to the translation model [35, 62, 69, 73, 84, 85, 90, 94, 97], or they can augment other spatial or spatio-temporal features [36, 79, 80, 88, 89, 95]. Often, the keypoints are used as a sign language representation directly. In other cases they are processed using a graph neural network to map them onto an embedding space before translation [89, 94].

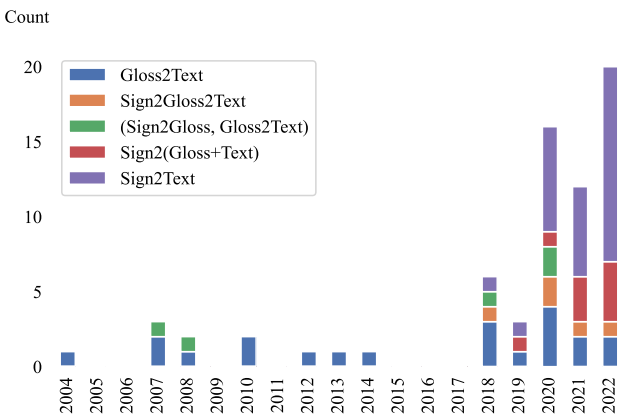


Fig. 6 Gloss-based models are used throughout the entire considered time period (2004–2022), but since 2018 models which translate from video to text are gaining traction

[5] As one paper may discuss several tasks, the total count is higher than the amount of papers.

Ten papers use 3D CNNs for feature extraction [35, 66, 67, 76, 82, 83, 86, 88, 96, 99]. These networks are able to extract spatio-temporal features, leveraging the temporal relations between neighboring frames in video data. The output of a 3D CNN is typically a sequence that is shorter than the input, summarizing multiple frames in a single feature vector. Similarly to 2D CNNs, these networks can be pre-trained on general tasks such as action recognition (on Kinetics [275]) or on more specific tasks such as isolated SLR (e.g., on WL-ASL [276]). Chen et al. [99] and Shi et al. [82] have shown independently that pre-training on sign language specific tasks yields better downstream SLT scores.

CNNs were state of the art in image feature extraction for several years. More recently, the vision transformer architecture was created that can outperform CNNs in certain scenarios [277]. Li et al. are the first to leverage vision transformers for feature extraction [89].

Kumar et al. opt to use traditional computer vision techniques instead of deep neural networks. They represent the hands and the face of the signer in the video as a set of contours [59]. First, they perform binarization to segment the hands and the face based on skin tone. Then they use the active contours method [278] to detect the edges of the hands and face. These are normalized with respect to the signer's position in the video frame by representing every coordinate as an angle (binned to 360 different angles).

5.4.2 Multi-cue approaches

A simple approach to feature extraction is to consider full video frames as inputs. Performing further pre-processing of the visual information to target hands, face and pose information separately (referred to as a multi-cue approach) improves the performance of SLT models [36, 59, 65, 75, 80, 86, 96]. Zheng et al. [75] show through qualitative analysis that adding facial feature extraction improves translation accuracy in utterances where facial expressions are used. Dey et al. [96] observe improvements in BLEU scores when adding lip reading as an input channel. By adding face crops as an additional channel, Miranda et al. [86] improve the performance of the TSPNet architecture [66].

5.5 Sign language translation models

The current SOTA in SLT is entirely based on encoder-decoder NMT models. RNNs are evaluated in 16 papers [6, 35, 59–62, 64, 65, 67, 70, 75, 76, 79, 80, 91, 94] and transformers in 34 papers [35–38, 60, 62, 65, 66, 68, 69, 71–74, 77, 78, 81–93, 95, 96, 98–100]. Within the RNN-based models, several attention schemes are used: no attention, Luong attention [279] and Bahdanau attention [42].

To the best of our knowledge, there has been no systematic comparison of RNNs and transformers across multiple tasks and datasets for SLT. Some authors perform a comparison between both architectures on specific datasets with specific sign language representations. A conclusive meta-study across papers is problematic due to inter-paper differences.

Ko et al. [62] report that RNNs with Luong attention obtain the highest ROUGE score, but transformers perform better in terms of METEOR, BLEU, and CIDEr (on the KETI dataset). In their experiments, Luong attention outperforms Bahdanau attention and RNNs without attention.

Moe et al. [60] compare RNNs and transformers for Gloss2Text with different tokenization schemes, and in every one of the experiments (on their own dataset), the transformer outperforms the RNN.

Four papers compare RNNs and transformers on RWTH-PHOENIX-Weather 2014T. Orbay et al. [35] report that an RNN with Bahdanau attention outperforms both an RNN with Luong attention and a transformer in terms of ROUGE and BLEU scores. Yin et al. [65] find that a transformer outperforms RNNs and that an RNN with Luong attention outperforms one with Bahdanau attention. Angelova et al. [91] achieve higher scores with RNNs than with transformers (on the DGS corpus [280] as well). Finally, Camgöz et al. [37] report a large increase in BLEU scores when using transformers, compared to their previous paper using RNNs [6]. However, the comparison is between models with different feature extractors and the impact of the architecture versus that of the feature extractors is not evaluated. It is likely that replacing a 2D CNN pre-trained on ImageNet [274] image classification with one pre-trained on CSLR will result in a significant increase in performance, especially when the CSLR model was trained on data from the same source (i.e., RWTH-PHOENIX-Weather 2014), as is the case here.

Pre-trained language models are readily available for transformers (for example via the HuggingFace Transformers library [281]). De Coster et al. have shown that integrating pre-trained spoken language models can improve SLT performance [38, 92]. Chen et al. pre-train their decoder network in two steps: first on a multilingual corpus, and then on Gloss2Text translation [99]. This pre-training approach can drastically improve performance. Chen et al. outperform other models on the RWTH-PHOENIX-Weather 2014T dataset [99]: 28.39 BLEU-4 (the second highest score is 25.59).

5.6 Evaluation

The majority of evaluation studies of the quality of SLT models is based on quantitative metrics. Eight different metrics are used across the 57 papers: BLEU, ROUGE, WER, TER, PER, CIDEr, METEOR, COMET and NIST.

A total of 22 papers [5, 6, 37, 51, 52, 61, 63–66, 70, 72, 75, 77, 79, 81–83, 86, 92, 93, 97] also provide a small set of example translations, along with ground truth reference

Table 3 Performance of different models on RWTH-PHOENIX-Weather 2014T Gloss2Text translation

References	Year	Architecture	BLEU-4	ROUGE	METEOR	COMET
[6]	2018	RNN	19.26	45.45	–	–
[70]	2020	RNN	17	41.5	–	–
			16.7	40.7	–	–
			17	43.1	–	–
			18.1	43.5	–	–
			17.8	42.8	–	–
[37]	2020	Transformer	24.54	–	–	–
[65]	2020	Transformer	23.32	46.58	44.85	–
			24.9	48.51	46.25	–
[74]	2021	Transformer	22.02	–	–	6.84
			23.35	–	–	13.65
			23.17	–	–	11.7
[71]	2021	Transformer	24.38	–	–	–
[91]	2022	RNN	22.2	–	–	–
[91]	2022	Transformer	18.5	–	–	–

Table 4 Performance of different models on RWTH-PHOENIX-Weather 2014T Sign2Gloss2Text translation

References	Year	Representation	Architecture	BLEU-4	ROUGE	METEOR
[6]	2018	Spatial	RNN	18.13	43.8	–
[37]	2020	Spatial	Transformer	22.45	–	–
[65]	2020	Spatio-temporal	RNN	21.54	45.5	44.87
		multi-cue		21.75	45.66	44.84
			Transformer	24	46.77	45.78
				25.4	48.78	47.6
[77]	2021	Spatio-temporal	Transformer	23.51	49.35	–
[94]	2022	Spatio-temporal	RNN	22.3	–	–

translations, allowing for qualitative analysis. Dreuw et al.'s model outputs mostly correct translations, but with different word order than the ground truth [51]. Camgöz et al. mention that the most common errors are related to numbers, dates, and places: these can be difficult to derive from context in weather broadcasts [6, 37]. The same kind of errors is made by the models of Partaourides et al. [70] and Voskou et al. [81]. Zheng et al. illustrate how their model improves accuracy for longer sentences [64]. Including facial expressions in the input space improves the detection of emphasis laid on adjectives [75].

The datasets used in the WMTSLT22 task, FocusNews and SRF, have a broader domain (news broadcasts) than, e.g., the RWTH-PHOENIX-Weather 2014T dataset (weather broadcasts). This makes the task significantly more challenging, as can be observed in the range of BLEU scores that are achieved (typically less than 1, compared to scores in the twenties for RWTH-PHOENIX-Weather 2014T). Example translation outputs also provide insight here. The models of Tarres et al. [97] and Hamidullah et al. [83] simply predict the most common German words in many cases, indicating that the SLT model has failed to learn the structure of the data. Shi's model [82] only translates phrases correctly when they occur in the training set, suggesting overfitting. Angelova et al. use the DGS corpus [280] (which contains discourse on general topics) as a dataset; they also obtain much lower translation scores than on RWTH-PHOENIX-Weather 2014T [91].

To the best of our knowledge, none of the papers discussed in this overview contain evaluations by members of SLCs. Two papers perform human evaluation, but only by hearing people. Luqman et al. [63] ask native Arabic speakers to evaluate the model's output translations on a three-point scale. For the WMTSLT22 challenge [85], translation outputs were scored by human evaluators (native German speakers trained as DSGS interpreters). The resulting scores indicate a considerable gap between the performance of human translators (87%) and MT (2%).

5.7 The RWTH-PHOENIX-Weather 2014T benchmark

The popularity of the RWTH-PHOENIX-Weather 2014T dataset facilitates the comparison of different SLT models on this dataset. We compare models based on their BLEU-4

Table 5 Performance of different models on RWTH-PHOENIX-Weather 2014T (Sign2Gloss,Gloss2Text) translation

References	Year	Representation	Architecture	BLEU-4	ROUGE	METEOR
[6]	2018	Spatial	RNN	17.79	43.45	–
[37]	2020	Spatial	Transformer	21.59	–	–
[65]	2020	Spatio-temporal multi-cue	Transformer	23.77	47.32	45.54

Table 6 Performance of different models on RWTH-PHOENIX-Weather 2014T Sign2(Gloss+Text) translation

References	Year	Representation	Architecture	BLEU-4	ROUGE
[37]	2020	Spatial	Transformer	21.32	–
[80]	2021	Spatio-temporal, multi-cue	RNN	23.65	46.65
[38]	2021	Spatial	Transformer	22.25	–
				21.16	–
				16.64	–
[78]	2021	Spatial	Transformer	23.14	49.23
[99]	2022	Spatio-temporal	Transformer	28.39	52.65
[92]	2022	Spatial	Transformer	21.82	47.25
[89]	2022	Spatio-temporal	Transformer	22.52	–
[88]	2022	Spatio-temporal	Transformer	19.3	–

score as this is the only metric consistently reported on in all of the papers using RWTH-PHOENIX-Weather 2014T (except [86]).

An overview of Gloss2Text models is shown in Table 3. For Sign2Gloss2Text, we refer to Table 4, and for (Sign-2Gloss, Gloss2Text) to Table 5. For Sign2(Gloss+Text) and Sign2Text, we list the results in Tables 6 and 7, respectively.

5.7.1 Sign language representations

Six papers use features extracted using a 2D CNNs by first training a CSLR model on RWTH-PHOENIX-Weather 2014[6] [6, 36, 38, 81, 92, 93]. These papers use the full frame as inputs to the feature extractor.

Others combine multiple input channels. Yin et al. [65] use Spatio-Temporal Multi-Cue (STMC) features, extracting images of the face, hands and full frames as well as including estimated poses of the body. These features are processed by a network which performs temporal processing, both on the intra- and the inter-cue level. Their models are the SOTA of Sign2Gloss2Text translation (25.4 BLEU-4). The model by Zhou et al. is similar and obtains a BLEU-4 score of 23.65 on Sign2(Gloss+Text) translation [80]. Camgöz et al. [36] use mouth pattern cues, pose information and hand shape information; by using this multi-cue representation, they are able to remove glosses from their translation model (but their feature extractors are still trained using glosses). Zheng et al. [75] use an additional channel of facial information for Sign2Text and obtain an increase of 1.6 BLEU-4 compared to their baseline. Miranda et al. [86] augment TSPNet [66] with face crops, improving the performance of the network.

Frame-based feature representations result in long input sequences to the translation model. The length of these sequences can be reduced by considering short clips instead of frames. This is done by using a pre-trained 3D CNN or by reducing the sequence length using temporal convolutions or RNNs that are trained jointly with the translation model. Zhou et al. [77] use 2D CNN features extracted from full frames, which are then further processed using temporal convolutions, reducing the temporal feature size by a factor 4. They call this approach Temporal Inception Networks (TIN). They achieve near-SOTA performance on Sign2Gloss2Text translation (23.51 BLEU-4) and Sign-2Text translation (24.32 BLEU-4). Zheng et al. [64] use an unsupervised algorithm called Frame Stream Density Compression (FSDC) to remove temporally redundant frames by comparing frames on the level of pixels. The resulting features are processed using a combination of temporal convolutions and RNNs. They compare the different settings and their combination and find that these techniques can be used to reduce the input size of the sign language features and to increase the BLEU-4 score. Chen et al. [99] achieve SOTA results of 28.39 BLEU-4 using 3D CNNs pre-trained first on Kinetics-400 [275] and then on WL-ASL [276].

[6] As discussed in Sect. 5.2, this dataset is an earlier version of RWTH-PHOENIX-Weather 2014T and they contain the same videos.

Table 7 Performance of different models on RWTH-PHOENIX-Weather 2014T Sign2Text translation

References	Year	Representation	Architecture	BLEU-4	ROUGE
[6]	2018	Spatial	RNN	9.58	31.8
[37]	2020	Spatial	Transformer	20.17	–
[36]	2020	Spatial, multi-cue	Transformer	19.21	45.05
				18.51	43.57
[35]	2020	Spatial	RNN	9.4	29.41
				9.33	29.62
				9.06	29.09
		Spatio-temporal		8.76	29.74
				8.26	28.64
				8.09	28
		Spatial (pose)		10.92	32.85
				10.23	31.47
				9.91	30.65
		Spatial		12.17	34.59
				11.15	31.98
				12.21	34.41
				13.25	36.28
[64]	2020	Spatial	RNN	9.76	31.34
				12.4	31.2
				10.66	32.25
				9.71	31.52
				10.73	32.99
[66]	2020	Spatio-temporal	Transformer	12.97	34.77
				13.41	34.95
[72]	2021	Spatial	Transformer	15.18	38.85
[76]	2021	Spatio-temporal	RNN	4.56	–
[77]	2021	Spatio-temporal	Transformer	24.32	49.54
[81]	2021	Spatial	Transformer	25.59	–
[75]	2021	Spatial, multi-cue	RNN	10.89	34.88
[79]	2021	Spatial, spatio-temporal (pose)	RNN	24.8	54.8
[96]	2022	Spatial	Transformer	20.24	–
[96]	2022	Spatio-temporal	Transformer	13.22	–
[95]	2022	Spatial	Transformer	12.34	–
[92]	2022	Spatial	Transformer	21.39	46.67
[87]	2022	Spatial	Transformer	24.02	49.97
[86]	2022	Spatio-temporal	Transformer	–	35.58

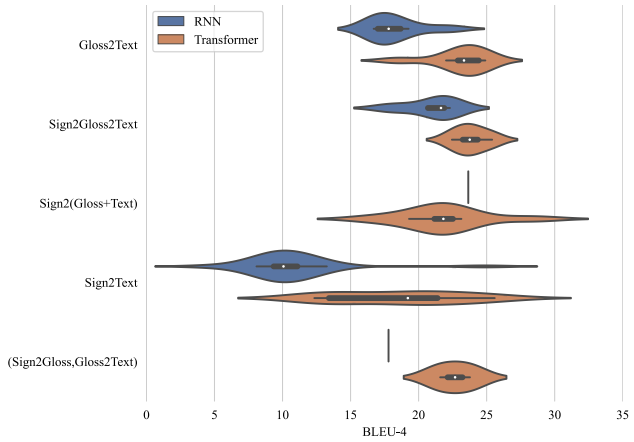


Fig. 7 Transformers tend to outperform RNNs on different SLT tasks in terms of BLEU-4 score on the RWTH-PHOENIX-Weather 2014T dataset

5.7.2 Neural architectures

We investigate whether RNNs or transformers perform best on this dataset. As this may depend on the used sign language representation, we analyze Gloss2Text, Sign2Gloss-2Text, (Sign2Gloss,Gloss2Text) Sign2(Gloss+Text) and Sign2Text separately.

Because all Gloss2Text models use the same sign language representation (glosses), we can directly compare the performance of different encoder-decoder architectures. Transformers (23.02 ± 2.05) outperform RNNs (18.29 ± 1.931).

The Sign2Gloss2Text transformer models by Yin et al. [65] achieve better performance (23.84 ± 1.225) than their recurrent models (20.47 ± 2.032).

There is only a single (Sign2Gloss, Gloss2Text) model using an RNN, and it achieves 17.79 BLEU-4 [6]. The transformer models of Camgöz et al. [37] and Yin et al. [65] achieve 21.59 and 23.77, respectively. These models all use different feature extractors, so direct comparison is not possible.

No direct comparison is available for Sign2(Gloss+Text) translation. Zhou et al. [80] present an LSTMs encoder-decoder using spatio-temporal multi-cue features and obtain 24.32 BLEU-4. The best Sign2(Gloss+Text) model leverages the pre-trained large language model mBART [282] (a transformer) and obtains 28.39 BLEU-4 [283].

The Sign2Text translation models exhibit higher variance in their scores than models for the other tasks. This is likely due to the lack of additional supervision signal in the form of glosses: the choice of sign language representation has a larger impact on the translation score. The difference in BLEU-4 score between transformers (18.51 ± 4.68) and RNNs (10.72 ± 3.63) is larger than in other tasks. However,

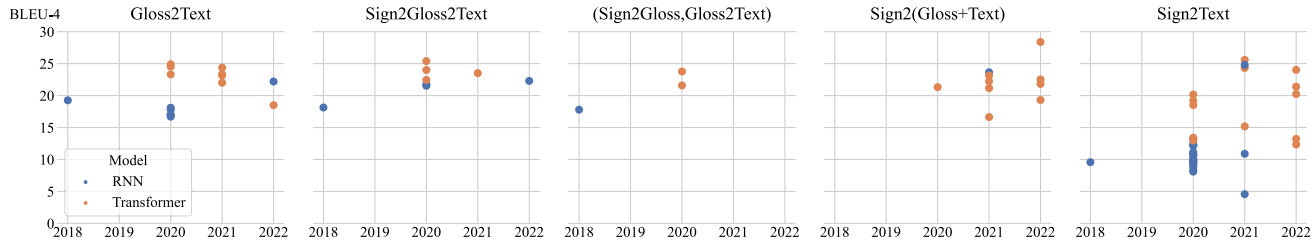


Fig. 8 Evolution of model scores on the RWTH-PHOENIX-Weather 2014T per task

we do not draw definitive conclusions from these results, as the sign language representations differ in architecture and pre-training task between models.

We provide a graphical overview of the performance of RNNs and transformers across tasks in Fig. 7 and observe that transformers often outperform RNNs on RWTH-PHOENIX-Weather 2014T. However, we cannot conclusively state whether this is due to the network architecture, or due to the sign language representations that these models are trained with.

5.7.3 Evolution of scores

Figure 8 shows an overview of the BLEU-4 scores on the RWTH-PHOENIX-Weather 2014T dataset from 2018 until 2023. It illustrates that the current best performing model (28.39 BLEU-4) is a Sign2(Gloss+Text) transformer proposed by Chen et al. [99].

6 Discussion of the current state of the art

The analysis of the scientific literature on SLT in Sect. 5 allows us to formulate answers to the four research questions.

6.1 RQ1: Datasets

RQ1 asks, *"Which datasets are used and what are their properties?"* The most frequently used dataset is RWTH-PHOENIX-Weather 2014T [6] for translation from DGS to German. It contains 8257 parallel utterances from several different interpreters. The domain is weather broadcasts.

Current datasets have several limitations. They are typically restricted to controlled domains of discourse (e.g., weather broadcasts) and have little variability in terms of visual conditions (e.g., TV studios). Camgöz et al. recently introduced three new benchmark datasets from the TV news and weather broadcasts domain [73]. Two similar datasets were introduced in 2022 by Müller et al. [85]. Because news broadcasts are included, the domain of discourse (and thus the vocabulary) is broader. It is more challenging to achieve acceptable translation performance with broader domains [56, 73, 83, 91, 97]. Yet, these datasets are more representative of real-world signing.

Another limitation is not related to the content, but rather the style of signing. Many SLT datasets contain recordings of non-native signers. In several cases, the signing is interpreted (often under time pressure) from spoken language. This means that the used signing may not be representative of the sign language and may in fact be influenced by the grammar of a spoken language. Training a translation model on these kinds of data has implications for the quality and accuracy of the resulting translations.

6.2 RQ2: Sign language representations

RQ2 asks, *"Which kinds of sign language representations are most informative?"* The limitations and drawbacks of glosses lead to the use of visual-based sign language representations. This representation can have a large impact on the performance of the SLT model. Spatio-temporal and multi-cue sign language representations outperform simple spatial (frame-based) sign language representations. Pre-training on SLR tasks yields better features for SLT.

6.3 RQ3: Translation model architectures

RQ3 asks, *"Which algorithms are currently the SOTA for SLT?"* Despite the generally small size of the datasets used for SLT, we see that neural MT models achieve the highest translation scores. Transformers outperform RNNs in many cases, but our literature overview suggests that the choice of sign language representation has a larger impact than the choice of translation architecture.

6.4 RQ4: Evaluation

RQ4 asks, *"How are current SLT models evaluated?"* Many papers report several translation related metrics, such as BLEU, ROUGE, WER and METEOR. These are standard metrics in MT. Several papers also provide example translations to allow the reader to gauge the translation quality for themselves. Whereas the above metrics often correlate quite well with human evaluation, this is not always the case

[284]. They also sometimes do not correlate among each other (what is the best model can be different depending on the considered metric). Only two of the 57 reviewed papers incorporate human evaluators in the loop [63, 85]. None of the reviewed papers evaluate their models in collaboration with native signers.

7 Challenges and proposals

Our literature overview (Sect. 5) and discussion thereof (Sect. 6) illustrate that the current challenges in the domain are threefold: (i) the collection of datasets, (ii) the design of sign language representations, and (iii) evaluation of the proposed models. We discuss these below, and finally give suggestions for the development of SLT models with SOTA methods.

7.1 Dataset collection

7.1.1 Challenges

Currently, SLT is a low-resource MT task: the largest public video datasets for MT contain just thousands of training examples (see Table 2). Current research uses datasets in which the videos have fixed viewpoints, similar backgrounds, and sometimes the signers even wear similar clothing for maximum contrast with the background. Yet, in real-world applications, dynamic viewpoints and lighting conditions will be a common occurrence. Furthermore, far from all sign languages have corresponding translation datasets. Additional datasets need to be collected and existing ones need to be extended.

Current datasets are insufficiently large to support SLT on general topics. When moving from weather broadcasts to news broadcasts, we observe a significant drop in translation scores. There is a clear trade-off between dataset size and vocabulary size.

De Meulder [285] raises concerns with current dataset collection efforts. Existing datasets and those currently being collected suffer from several biases. If interpreted data are used, influence from spoken languages will be present in the dataset. If only native signer data are used, then the majority of signers will have the same ethnicity. Both statistical as well as neural MT exacerbate bias [286, 287]. Therefore, when our training datasets are biased and of small volumes, we cannot expect (data driven) MT systems to reach high qualities and be generalizable.

7.1.2 Proposals

We propose to gather two kinds of datasets: focused datasets for training SLT models, but also large, multi-lingual datasets for the design of sign language representations. The former type of datasets already exists, but the latter kind, to the best of our knowledge, does not yet exist.

By collecting larger, multilingual datasets, we can learn sign language representations with (self-)supervised deep learning techniques. Such datasets do not need to consist entirely of native signing. They should include many topics and visual characteristics to be as general as possible.

In contrast, SLT requires high quality labeled data, the collection of which is challenging. Bragg et al.'s first and second calls to action, "Involve Deaf team members throughout" and "Focus on real-world applications" [17], guide the dataset collection process. By involving SLC members, the dataset collection effort can be guided toward use cases that would benefit SLCs. Additionally, by collecting datasets with a limited domain of discourse targeted at specific use cases, the SLT problem is effectively simplified. As a result, any applications would be limited in scope, but more useful in practice.

7.2 Sign language representations

7.2.1 Challenges

Current sign language representations do not take into account the productive lexicon. In fact, it is doubtful whether a pure end-to-end NMT approach is capable of tackling productive signs. To recognize and understand productive signs, we need models that have the ability to link abstract visual information to the properties of objects. Incorporating the productive lexicon in translation systems is a significant challenge, one for which, to the best of our knowledge, labeled data is currently not available.

Current end-to-end representations moreover do not explicitly account for fingerspelling, signing space, or classifiers. Learning these aspects in the translation model with an end-to-end approach is challenging, especially due the scarcity of annotated data.

Our literature overview shows that the choice of representation has a significant impact on the translation performance. Hence, improving the feature extraction and incorporating the aforementioned sign language characteristics is paramount.

7.2.2 Proposals

Linguistic analysis of sign languages can inform the design of sign language representations. The definition of the so-called meaningful units has been discussed by De Sisto et al. [32]. It requires collaboration between computer scientists and (computational) linguists. Researchers should analyze the representations that are automatically learned by SOTA SLT models. For example, SLR models appear to implicitly

learn to recognize hand shapes [288]. Based on such analyses, linguists can suggest which components to focus on next.

In parallel, we can exploit unlabeled sign language data to learn sign language representations in a self-supervised manner. Recently, increasingly larger neural networks are being trained on unlabeled datasets to discover latent patterns and to learn neural representations of textual, auditory, and visual data. In the domain of natural language processing, we already observe tremendous advances thanks to self-supervised language models such as BERT [45]. In computer vision, self-supervised techniques are applied to pre-train powerful feature extractors which can then be applied to downstream tasks such as image classification or object detection. Algorithms such as SimCLR [283], BYOL [289] and DINO [290] are used to train 2D CNNs and vision transformers without labels, reaching performance that is almost on the same level as models trained with supervised techniques. In the audio domain, Wav2Vec 2.0 learns discrete speech units in a self-supervised manner [291]. In sign language processing, self-supervised learning can be applied to train spatio-temporal representations (like Wav2Vec 2.0 or SimCLR), and to contextualize those representations (like BERT).

Sign languages share some common elements, for example the fact that they all use the human body to convey information. Movements used in signing are composed of motion primitives and the configuration of the hand (shape and orientation) is important in all sign languages. The recognition of these low level components does not require language specific datasets and could be performed on multilingual datasets, containing videos recorded around the world with people of various ages, genders, and ethnicities. The representations extracted from multilingual SLR models can then be fine-tuned in monolingual or multilingual SLT models.

Self-supervised and multilingual learning should be evaluated for the purpose of learning such common elements of sign languages. This will not only facilitate automatic SLT, but could also lead to the development of new tools supporting linguistic analysis of sign languages and their commonalities and differences.

7.3 Evaluation

7.3.1 Challenges

Current research uses mostly quantitative metrics to evaluate SLT models, on datasets with limited scope. In-depth error analysis is missing from many SLT papers. SLT models should also be evaluated on real-world data from real-world settings. Furthermore, human evaluation from signers and non-signers is required to truly assess the translation quality. This is especially true because many of the SLT models are currently designed, implemented and evaluated by hearing researchers.

7.3.2 Proposals

Human-in-the-loop development can alleviate some of the concerns that live in SLCs about the application of MT techniques to sign languages about appropriation of sign languages. Human (signing and non-signing) evaluators should be included in every step of SLT research. Their feedback should guide the development of new models. For example, if the current models fail to properly translate classifiers, then SLT researchers could choose to focus on classifiers. This would hasten the progress in this field which is currently mostly focusing on improving metrics that say little about the usability of the SLT models.

Inspiration for human evaluation can be found in the yearly conference on machine translation (WMT), where researchers perform both direct assessment of translations, and relative ranking [292]. Müller et al. performed human evaluation on a benchmark dataset after an SLT challenge [85]. They hired native German speakers trained as DSGS interpreters to evaluate four different models, and compare their outputs to human translations. Their work can be a guideline for human evaluation in future research.

7.4 Applying SOTA techniques

There is still a large gap between MT and human level performance for SLT [85]. However, with the current SOTA and sufficient constraints, it may be possible to develop limited SLT applications. The development of these applications can be guided with the following three principles.

First, a dataset should be collected that has a specific topic related to the application: it is not yet possible to train robust SLT models with large vocabularies [56, 73, 83, 91, 97]. Second, the feature extractor should be pre-trained on SLR tasks as this yields the most informative representations [82, 96, 99]. Third, qualitative evaluation and evaluation by humans can provide insights into the failure cases of SLT models.

8 Conclusion

In this article, we discuss the SOTA of SLT and explore challenges and opportunities for future research through a systematic overview of the papers in this domain. We review 57 papers on machine translation from sign to spoken languages. These papers are selected based on predefined criteria and they are indicative of sound SLT research. The selected papers are written in English and peer-reviewed. They propose, implement and evaluate a

sign language machine translation system from a sign language to a spoken language, supporting RGB video inputs. We discuss the SOTA of SLT and explore several challenges and opportunities for future research.

In recent years, neural machine translation has become dominant in the growing domain of SLT. The most powerful sign language representations are those that combine information from multiple channels (manual actions, body movements and mouth patterns) and those that are reduced in length by temporal processing modules. These translation models are typically RNNs or transformers. Transformers outperform RNNs in many cases, and large language models allow for transfer learning. SLT datasets are small: we are dealing with a low-resource machine translation problem. Many datasets consider limited domains of discourse and generally contain recordings of non-native signers. This has implications on the quality and accuracy of translations generated by models trained on these datasets, which must be taken into account when evaluating SLT models. Datasets that consider a broader domain of discourse are too small to train NMT models on. Evaluation is mostly performed using quantitative metrics that can be computed automatically, given a corpus. There are currently no works that perform evaluation of neural SLT models in collaboration with sign language users.

Progressing beyond the current SOTA of SLT requires efforts in data collection, the design of sign language representations, machine translation, and evaluation. Future research may improve sign language representations by incorporating domain knowledge into their design and by leveraging abundant, but as of yet unexploited, unlabeled data. Research should be conducted in an interdisciplinary manner, with computer scientists, sign language linguists, and experts on sign language cultures working together. Finally, SLT models should be evaluated in collaboration with end users: native signers as well as hearing people that do not know any sign language.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10209-023-00992-1.

Acknowledgments The authors wish to thank the anonymous reviewers for their detailed evaluation of our manuscript.

Funding Mathieu De Coster's research is funded by the Research Foundation Flanders (FWO Vlaanderen): file number 77410. This work has been conducted within the SignON project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017255.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Pugeault, N., Bowden, R.: Spelling it out: Real-time asl fingerspelling recognition. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1114–1119 (2011). IEEE
2. Fowley, F., Ventresque, A.: Sign language fingerspelling recognition using synthetic data. In: AICS, pp. 84–95 (2021). CEUR-WS
3. Pigou, L., Van Herreweghe, M., Dambre, J.: Gesture and sign language recognition with temporal residual networks. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 3086–3093 (2017)
4. Jiang, S., Sun, B., Wang, L., Bai, Y., Li, K., Fu, Y.: Skeleton aware multi-modal sign language recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3413–3423 (2021)
5. Bungeroth, J., Ney, H.: Statistical sign language translation. In: Workshop on Representation and Processing of Sign Languages, LREC, vol. 4, pp. 105–108 (2004). Citeseer
6. Camgoz, N.C., Hadfield, S., Koller, O., Ney, H., Bowden, R.: Neural sign language translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7784–7793 (2018). https://doi.org/10.1109/CVPR.2018.00812
7. Stein, D., Bungeroth, J., Ney, H.: Morpho-syntax based statistical methods for automatic sign language translation. In: Proceedings of the 11th Annual Conference of the European Association for Machine Translation (2006)
8. Morrissey, S., Way, A.: Joining hands: Developing a sign language machine translation system with and for the deaf community. In: CVHI (2007)
9. San-Segundo, R., López, V., Martın, R., Sánchez, D., Garcıa, A.: Language resources for spanish–spanish sign language (lse) translation. In: Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies at LREC, pp. 208–211 (2010)
10. David, B., Bouillon, P.: Prototype of automatic translation to the sign language of french-speaking Belgium. Evaluation by the deaf community. Modelling, Measurement and Control C 79(4), 162–167 (2018)
11. Erard, M.: Why sign language gloves don't help deaf people (2017). https://www.theatlantic.com/technology/archive/2017/11/why-sign-language-gloves-dont-help-deaf-people/545441/
12. Adnan, N.H., Wan, K., AB, S., BAKAR, J.A.A.: Learning and manipulating human's fingertip bending data for sign language translation using pca-bmu classifier. CREAM: Curr. Res. Malaysia (Penyelidikan Terkini di Malaysia) **3**, 361–372 (2013)
13. Caliwag, A., Angsanto, S.R., Lim, W.: Korean sign language translation using machine learning. In: 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 826–828 (2018). IEEE

14. Mistry, J., Inden, B.: An approach to sign language translation using the intel realsense camera. In: 2018 10th Computer Science and Electronic Engineering (CEEC), pp. 219–224 (2018). IEEE
15. Krishnan, P.T., Balasubramanian, P.: Detection of alphabets for machine translation of sign language using deep neural net. In: 2019 International Conference on Data Science and Communication (IconDSC), pp. 1–3 (2019). IEEE
16. Núñez-Marcos, A., Perez-de-Viñaspre, O., Labaka, G.: A survey on sign language machine translation. Expert Systems with Applications, 118993 (2022)
17. Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., *et al.*: Sign language recognition, generation, and translation: An interdisciplinary perspective. In: The 21st International ACM SIGACCESS Conference on Computers and Accessibility, pp. 16–31 (2019)
18. Vermeerbergen, M., Twilhaar, J.N., Van Herreweghe, M.: Variation between and within sign language of the netherlands and flemish sign language. In: Language and Space Volume 30 (3): Dutch, pp. 680–699. De Gruyter Mouton, Berlin (2013)
19. Van Herreweghe, M., Vermeerbergen, M.: Flemish sign language standardisation. Current issues in language planning **10**(3), 308–326 (2009)
20. Stokoe, W.: Sign language structure: An outline of the visual communication systems of the american deaf. Studies in Linguistics, Occasional Papers **8** (1960)
21. Battison, R.: Lexical Borrowing in American Sign Language. Linstok Press, Silver Spring (1978)
22. Bank, R., Crasborn, O.A., Van Hout, R.: Variation in mouth actions with manual signs in Sign Language of the Netherlands (NGT). Sign Language & Linguistics **14**(2), 248–270 (2011)
23. Perniss, P.: 19. use of sign space. In: Sign Language, pp. 412–431. De Gruyter Mouton, Berlin (2012)
24. Zwitserlood, I.: In: Pfau, R., Steinbach, M., Woll, B. (eds.) Classifiers, pp. 158–186. De Gruyter Mouton, Berlin (2012). https://doi.org/10.1515/9783110261325.158
25. Vermeerbergen, M.: Past and current trends in sign language research. Lang. Commun. **26**(2), 168–192 (2006). https://doi.org/10.1016/j.langcom.2005.10.004
26. Frishberg, N., Hoiting, N., Slobin, D.I.: In: Pfau, R., Steinbach, M., Woll, B. (eds.) Transcription, pp. 1045–1075. De Gruyter Mouton, Berlin (2012). https://doi.org/10.1515/9783110261325.1045
27. Sutton, V.: Sign Writing for Everyday Use. Sutton Movement Writing Press, New York (1981)
28. Prillwitz, S.: HamNoSys Version 2.0. Hamburg Notation System for Sign Languages: An Introductory Guide. Intern. Arb. z. Gebärdensprache u. Kommunik. Signum Press, Berlin (1989)
29. Vermeerbergen, M., Leeson, L., Crasborn, O.A.: Simultaneity in Signed Languages: Form and Function vol. 281. John Benjamins Publishing, Amsterdam (2007). https://doi.org/10.1075/cilt.281
30. De Sisto, M., Vandeghinste, V., Gómez, S.E., De Coster, M., Shterionov, D., Seggion, H.: Challenges with sign language datasets for sign language recognition and translation. In: LREC2022, the 13th International Conference on Language Resources and Evaluation, pp. 2478–2487 (2022)
31. Murtagh, I.E.: A linguistically motivated computational framework for irish sign language. PhD thesis, Trinity College Dublin. School of Linguistic Speech and Comm Sci (2019)
32. De Sisto, M., Shterionov, D., Murtagh, I., Vermeerbergen, M., Leeson, L.: Defining meaningful units. challenges in sign segmentation and segment-meaning mapping. In: Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pp. 98–103. Association for Machine Translation in the Americas, Virtual (2021). https://aclanthology.org/2021.mtsummit-at4ssl.11
33. Koller, O., Camgoz, N.C., Ney, H., Bowden, R.: Weakly supervised learning with multi-stream cnn-lstm-hmms to discover sequential parallelism in sign language videos. IEEE Trans. Pattern Anal. Mach. Intell. **42**(9), 2306–2320 (2019). https://doi.org/10.1109/TPAMI.2019.2911077
34. Koller, O., Ney, H., Bowden, R.: Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3793–3802 (2016)
35. Orbay, A., Akarun, L.: Neural sign language translation by learning tokenization. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), pp. 222–228 (2020). IEEE
36. Camgoz, N.C., Koller, O., Hadfield, S., Bowden, R.: Multi-channel transformers for multi-articulatory sign language translation. In: European Conference on Computer Vision, pp. 301–319 (2020). Springer
37. Camgoz, N.C., Koller, O., Hadfield, S., Bowden, R.: 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, pp. 10023–10033 (2020)
38. De Coster, M., D'Oosterlinck, K., Pizurica, M., Rabaey, P., Verlinden, S., Van Herreweghe, M., Dambre, J.: Frozen pretrained transformers for neural sign language translation. In: Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pp. 88–97. Association for Machine Translation in the Americas, Virtual (2021). https://aclanthology.org/2021.mtsummit-at4ssl.10
39. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)
40. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
41. Cho, K., van Merriënboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, pp. 1724–1734 (2014)
42. Bahdanau, D., Cho, K.H., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations, ICLR 2015 (2015)
43. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
44. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014). https://doi.org/10.3115/v1/D14-1162
45. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (2019). https://doi.org/10.18653/v1/N19-1423
46. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Jurafsky, D., Chai, J.,

Schluter, N., Tetreault, J.R. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL, pp. 7871–7880 (2020). https://doi.org/10.18653/v1/2020.acl-main.703. ACL
47. Harris, Z.: Distributional structure. Word **10**(2–3), 146–162 (1954). https://doi.org/10.1007/978-94-009-8467-7_1
48. Firth, J.: A synopsis of linguistic theory 1930–1955. In: Studies in Linguistic Analysis. Philological Society, Oxford (1957). reprinted in Palmer, F. (ed.): R. Firth, Longman, Harlow (1968)
49. Stahlberg, F.: Neural machine translation: A review. Journal of Artificial Intelligence Research **69**, 343–418 (2020)
50. Stein, D., Dreuw, P., Ney, H., Morrissey, S., Way, A.: Hand in hand: automatic sign language to English translation. In: Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages: Papers, Skövde, Sweden (2007). https://aclanthology.org/2007.tmi-papers.26
51. Dreuw, P., Stein, D., Ney, H.: Enhancing a sign language translation system with vision-based features. In: International Gesture Workshop, pp. 108–113 (2007). Springer
52. Morrissey, S., Way, A., Stein, D., Bungeroth, J., Ney, H.: Combining data-driven mt systems for improved sign language translation. In: European Association for Machine Translation (2007)
53. Dreuw, P., Stein, D., Deselaers, T., Rybach, D., Zahedi, M., Bungeroth, J., Ney, H.: Spoken language processing techniques for sign language recognition and translation. Technol. Disabil. **20**(2), 121–133 (2008)
54. López, V., San-Segundo, R., Martín, R., Lucas, J.M., Echeverry, J.D.: Spanish generation from spanish sign language using a phrase-based translation system. technology 9, 10 (2010)
55. Stein, D., Schmidt, C., Ney, H.: Sign language machine translation overkill. In: International Workshop on Spoken Language Translation (IWSLT) 2010 (2010)
56. Stein, D., Schmidt, C., Ney, H.: Analysis, preparation, and optimization of statistical sign language machine translation. Mach. Transl. **26**(4), 325–357 (2012)
57. Schmidt, C., Koller, O., Ney, H., Hoyoux, T., Piater, J.: Using viseme recognition to improve a sign language translation system. In: International Workshop on Spoken Language Translation, pp. 197–203 (2013). Citeseer
58. Forster, J., Schmidt, C., Koller, O., Bellgardt, M., Ney, H.: Extensions of the sign language recognition and translation corpus rwth-phoenix-weather. In: LREC, pp. 1911–1916 (2014)
59. Kumar, S.S., Wangyal, T., Saboo, V., Srinath, R.: Time series neural networks for real time sign language translation. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 243–248 (2018). https://doi.org/10.1109/ICMLA.2018.00043. IEEE
60. Moe, S.Z., Thu, Y.K., Thant, H.A., Min, N.W.: Neural machine translation between myanmar sign language and myanmar written text. In: the Second Regional Conference on Optical Character Recognition and Natural Language Processing Technologies for ASEAN Languages, pp. 13–14 (2018)
61. Arvanitis, N., Constantinopoulos, C., Kosmopoulos, D.: Translation of sign language glosses to text using sequence-to-sequence attention models. In: 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 296–302 (2019). https://doi.org/10.1109/SITIS.2019.00056. IEEE
62. Ko, S.-K., Kim, C.J., Jung, H., Cho, C.: Neural sign language translation based on human keypoint estimation. Appl. Sci. **9**(13), 2683 (2019)
63. Luqman, H., Mahmoud, S.A.: A machine translation system from arabic sign language to arabic. Univ. Access Inf. Soc. **19**(4), 891–904 (2020). https://doi.org/10.1007/s10209-019-00695-6
64. Zheng, J., Zhao, Z., Chen, M., Chen, J., Wu, C., Chen, Y., Shi, X., Tong, Y.: An improved sign language translation model with explainable adaptations for processing long sign sentences. Computational Intelligence and Neuroscience **2020** (2020)
65. Yin, K., Read, J.: Better sign language translation with stmc-transformer. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 5975–5989 (2020). https://doi.org/10.18653/v1/2020.coling-main.525
66. Li, D., Xu, C., Yu, X., Zhang, K., Swift, B., Suominen, H., Li, H.: Tspnet: Hierarchical feature learning via temporal semantic pyramid for sign language translation. Adv. Neural. Inf. Process. Syst. **33**, 12034–12045 (2020)
67. Rodriguez, J., Chacon, J., Rangel, E., Guayacan, L., Hernandez, C., Hernandez, L., Martinez, F.: Understanding motion in sign language: A new structured translation dataset. In: Proceedings of the Asian Conference on Computer Vision (2020)
68. Moe, S.Z., Thu, Y.K., Thant, H.A., Min, N.W., Supnithi, T.: Unsupervised neural machine translation between myanmar sign language and myanmar language. tic 14(15), 16 (2020)
69. Kim, S., Kim, C.J., Park, H.-M., Jeong, Y., Jang, J.Y., Jung, H.: Robust keypoint normalization method for korean sign language translation using transformer. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1303–1305 (2020). https://doi.org/10.1109/ICTC49870.2020.9289551. IEEE
70. Partaourides, H., Voskou, A., Kosmopoulos, D., Chatzis, S., Metaxas, D.N.: Variational bayesian sequence-to-sequence networks for memory-efficient sign language translation. In: International Symposium on Visual Computing, pp. 251–262 (2020). Springer
71. Zhang, X., Duh, K.: Approaching sign language gloss translation as a low-resource machine translation task. In: Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pp. 60–70. Association for Machine Translation in the Americas, Virtual (2021). https://aclanthology.org/2021.mtsummit-at4ssl.7
72. Zhao, J., Qi, W., Zhou, W., Nan, D., Zhou, M., Li, H.: Conditional sentence generation and cross-modal reranking for sign language translation. IEEE Trans. Multimedia (2021). https://doi.org/10.1109/TMM.2021.3087006
73. Camgöz, N.C., Saunders, B., Rochette, G., Giovanelli, M., Inches, G., Nachtrab-Ribback, R., Bowden, R.: Content4all open research sign language translation datasets. In: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), pp. 1–5 (2021). https://doi.org/10.1109/FG52635.2021.9667087
74. Moryossef, A., Yin, K., Neubig, G., Goldberg, Y.: Data augmentation for sign language gloss translation. In: Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pp. 1–11. Association for Machine Translation in the Americas, Virtual (2021). https://aclanthology.org/2021.mtsummit-at4ssl.1
75. Zheng, J., Chen, Y., Wu, C., Shi, X., Kamal, S.M.: Enhancing neural sign language translation by highlighting the facial expression information. Neurocomputing **464**, 462–472 (2021)
76. Rodriguez, J., Martinez, F.: How important is motion in sign language translation? IET Comput. Vision **15**(3), 224–234 (2021)
77. Zhou, H., Zhou, W., Qi, W., Pu, J., Li, H.: Improving sign language translation with monolingual data by sign back-translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1316–1325 (2021)
78. Yin, A., Zhao, Z., Liu, J., Jin, W., Zhang, M., Zeng, X., He, X.: Simulslt: End-to-end simultaneous sign language translation. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 4118–4127 (2021)

79. Gan, S., Yin, Y., Jiang, Z., Xie, L., Lu, S.: Skeleton-aware neural sign language translation. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 4353–4361 (2021)
80. Zhou, H., Zhou, W., Zhou, Y., Li, H.: Spatial-temporal multi-cue network for sign language recognition and translation. IEEE Trans. Multimedia (2021). https://doi.org/10.1109/TMM.2021.3059098
81. Voskou, A., Panousis, K.P., Kosmopoulos, D., Metaxas, D.N., Chatzis, S.: Stochastic transformer networks with linear competing units: Application to end-to-end sl translation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 11946–11955 (2021)
82. Shi, B., Brentari, D., Shakhnarovich, G., Livescu, K.: Ttic's wmt-slt 22 sign language translation system. In: Proceedings of the Seventh Conference on Machine Translation, pp. 989–993. Association for Computational Linguistics, Abu Dhabi (2022). https://aclanthology.org/2022.wmt-1.96
83. Hamidullah, Y., van Genabith, J., España-Bonet, C.: Spatio-temporal sign language representation and translation. In: Proceedings of the Seventh Conference on Machine Translation, pp. 977–982. Association for Computational Linguistics, Abu Dhabi (2022). https://aclanthology.org/2022.wmt-1.94
84. Hufe, L., Avramidis, E.: Experimental machine translation of the swiss german sign language via 3d augmentation of body keypoints. In: Proceedings of the Seventh Conference on Machine Translation, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics (2022)
85. Müller, M., Ebling, S., Avramidis, E., Battisti, A., Berger, M., Bowden, R., Braffort, A., Cihan Camgöz, N., España-Bonet, C., Grundkiewicz, R., Jiang, Z., Koller, O., Moryossef, A., Perrollaz, R., Reinhard, S., Rios, A., Shterionov, D., Sidler-Miserez, S., Tissi, K., Van Landuyt, D.: Findings of the first wmt shared task on sign language translation (wmt-slt22). In: Proceedings of the Seventh Conference on Machine Translation, pp. 744–772. Association for Computational Linguistics, Abu Dhabi (2022). https://aclanthology.org/2022.wmt-1.71
86. Miranda, P.B., Casadei, V., Silva, E., Silva, J., Alves, M., Severo, M., Freitas, J.P.: Tspnet-hf: A hand/face tspnet method for sign language translation. In: Ibero-American Conference on Artificial Intelligence, pp. 305–316 (2022). Springer
87. Jin, T., Zhao, Z., Zhang, M., Zeng, X.: Prior knowledge and memory enriched transformer for sign language translation. In: Findings of the Association for Computational Linguistics: ACL 2022, pp. 3766–3775 (2022)
88. Chen, Y., Zuo, R., Wei, F., Wu, Y., Liu, S., Mak, B.: Two-stream network for sign language recognition and translation. arXiv preprint arXiv:2211.01367 (2022)
89. Li, R., Meng, L.: Sign language recognition and translation network based on multi-view data. Appl. Intell. **52**(13), 14624–14638 (2022)
90. Dal Bianco, P., Ríos, G., Ronchetti, F., Quiroga, F., Stanchi, O., Hasperué, W., Rosete, A.: Lsa-t: The first continuous argentinian sign language dataset for sign language translation. In: Ibero-American Conference on Artificial Intelligence, pp. 293–304 (2022). Springer
91. Angelova, G., Avramidis, E., Möller, S.: Using neural machine translation methods for sign language translation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pp. 273–284 (2022)
92. De Coster, M., Dambre, J.: Leveraging frozen pretrained written language models for neural sign language translation. Information **13**(5), 220 (2022)
93. Jin, T., Zhao, Z., Zhang, M., Zeng, X.: Mc-slt: Towards low-resource signer-adaptive sign language translation. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 4939–4947 (2022)
94. Kan, J., Hu, K., Hagenbuchner, M., Tsoi, A.C., Bennamoun, M., Wang, Z.: Sign language translation with hierarchical spatio-temporal graph neural network. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3367–3376 (2022)
95. Chaudhary, L., Ananthanarayana, T., Hoq, E., Nwogu, I.: Signnet ii: A transformer-based two-way sign language translation model. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
96. Dey, S., Pal, A., Chaabani, C., Koller, O.: Clean text and full-body transformer: Microsoft's submission to the wmt22 shared task on sign language translation. In: Proceedings of the Seventh Conference on Machine Translation, pp. 969–976. Association for Computational Linguistics, Abu Dhabi (2022). https://aclanthology.org/2022.wmt-1.93
97. Tarres, L., Gállego, G.I., Giro-i-Nieto, X., Torres, J.: Tackling low-resourced sign language translation: Upc at wmt-slt 22. In: Proceedings of the Seventh Conference on Machine Translation, pp. 994–1000. Association for Computational Linguistics, Abu Dhabi (2022). https://aclanthology.org/2022.wmt-1.97
98. Yin, A., Zhao, Z., Jin, W., Zhang, M., Zeng, X., He, X.: Mlslt: Towards multilingual sign language translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5109–5119 (2022)
99. Chen, Y., Wei, F., Sun, X., Wu, Z., Lin, S.: A simple multi-modality transfer learning baseline for sign language translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5120–5130 (2022)
100. Mohamed, A., Hefny, H., et al.: A deep learning approach for gloss sign language translation using transformer. Journal of Computing and Communication **1**(2), 1–8 (2022)
101. Fraiwan, M., Khasawneh, N., Ershedat, H., Al-Alali, I., Al-Kofahi, H.: A kinect-based system for arabic sign language to speech translation. Int. J. Comput. Appl. Technol. **52**(2–3), 117–126 (2015)
102. Jin, C.M., Omar, Z., Jaward, M.H.: A mobile application of american sign language translation via image processing algorithms. In: 2016 IEEE Region 10 Symposium (TENSYMP), pp. 104–109 (2016). IEEE
103. Patil, Y., Krishnadas, S., Kastwar, A., Kulkarni, S.: American and indian sign language translation using computer vision. In: International Conference on Business Management, Innovation & Sustainability (ICBMIS) (2020)
104. Makarov, I., Veldyaykin, N., Chertkov, M., Pokoev, A.: American and russian sign language dactyl recognition and text-2sign translation. In: International Conference on Analysis of Images, Social Networks and Texts, pp. 309–320 (2019). Springer
105. Bukhari, J., Rehman, M., Malik, S.I., Kamboh, A.M., Salman, A.: American sign language translation through sensory glove; signspeak. International Journal of u-and e-Service, Science and Technology **8**(1), 131–142 (2015)
106. Joshi, A., Sierra, H., Arzuaga, E.: American sign language translation using edge detection and cross correlation. In: 2017 IEEE Colombian Conference on Communications and Computing (COLCOM), pp. 1–6 (2017). IEEE
107. Rizwan, S.B., Khan, M.S.Z., Imran, M.: American sign language translation via smart wearable glove technology. In: 2019 International Symposium on Recent Advances in Electrical Engineering (RAEE), vol. 4, pp. 1–6 (2019). IEEE
108. Halawani, S.M., Zaitun, A.: An avatar based translation system from arabic speech to arabic sign language for deaf people. International Journal of Information Science and Education **2**(1), 13–20 (2012)
109. Anand, M.S., Kumaresan, A., Kumar, N.M.: An integrated two way isl (indian sign language) translation system–a new

approach. International Journal of Advanced Research in Computer Science **4**(1) (2013)
110. Kanwal, K., Abdullah, S., Ahmed, Y.B., Saher, Y., Jafri, A.R.: Assistive glove for pakistani sign language translation. In: 17th IEEE International Multi Topic Conference 2014, pp. 173–176 (2014). IEEE
111. Angona, T.M., Shaon, A.S., Niloy, K.T.R., Karim, T., Tasnim, Z., Reza, S.S., Mahbub, T.N.: Automated bangla sign language translation system for alphabets by means of mobilenet. Telkomnika **18**(3), 1292–1301 (2020)
112. Hoque, M.T., Rifat-Ut-Tauwab, M., Kabir, M.F., Sarker, F., Huda, M.N., Abdullah-Al-Mamun, K.: Automated bangla sign language translation system: Prospects, limitations and applications. In: 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), pp. 856–862 (2016). IEEE
113. Oliveira, T., Escudeiro, P., Escudeiro, N., Rocha, E., Barbosa, F.M.: Automatic sign language translation to improve communication. In: 2019 IEEE Global Engineering Education Conference (EDUCON), pp. 937–942 (2019). IEEE
114. Ayadi, K., ElHadj, Y.O., Ferchichi, A.: Automatic translation from arabic to arabic sign language: A review. In: 2018 JCCO Joint International Conference on ICT in Education and Training, International Conference on Computing in Arabic, and International Conference on Geocomputing (JCCO: TICET-ICCA-GECO), pp. 1–5 (2018). IEEE
115. Mohandes, M.: Automatic translation of arabic text to arabic sign language. AIML Journal **6**(4), 15–19 (2006)
116. Fernandes, L., Dalvi, P., Junnarkar, A., Bansode, M.: Convolutional neural network based bidirectional sign language translation system. In: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 769–775 (2020). IEEE
117. Dabwan, B.A.: Convolutional neural network-based sign language translation system. International Journal of Engineering, Science and Mathematics **9**(6), 47–57 (2020)
118. Martin, V.: Design and implementation of a system for automatic sign language translation. In: Future Access Enablers of Ubiquitous and Intelligent Infrastructures, pp. 307–313 (2015). Springer
119. Yudhana, A., Rahmawan, J., Negara, C.: Flex sensors and mpu6050 sensors responses on smart glove for sign language translation. In: IOP Conference Series: Materials Science and Engineering, vol. 403, p. 012032 (2018). IOP Publishing
120. Mohanty, S., Prasad, S., Sinha, T., Krupa, B.N.: German sign language translation using 3d hand pose estimation and deep learning. In: 2020 IEEE REGION 10 CONFERENCE (TENCON), pp. 773–778 (2020). IEEE
121. Pranatadesta, R.A., Suwardi, I.S.: Indonesian sign language (bisindo) translation system with orb for bilingual language. In: 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), pp. 502–505 (2019). IEEE
122. Prasad, P.K., Shibu, A.P., et al.: Intelligent human sign language translation using support vector machines classifier. IJRAR-International Journal of Research and Analytical Reviews (IJRAR) **5**(4), 461–466 (2018)
123. Bajpai, D., Mishra, V.: Low cost full duplex wireless glove for static and trajectory based american sign language translation to multimedia output. In: 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 646–652 (2016). IEEE
124. Yang, S., Cui, X., Guo, R., Zhang, Z., Sang, S., Zhang, H.: Piezoelectric sensor based on graphene-doped pvdf nanofibers for sign language translation. Beilstein J. Nanotechnol. **11**(1), 1655–1662 (2020)
125. Gamarra, J.E.M., Cubas, M.A.S., Silupú, J.D.S., Chirinos, C.E.C.: Prototype for peruvian sign language translation based on an artificial neural network approach. In: 2020 IEEE XXVII International Conference on Electronics, Electrical Engineering and Computing (INTERCON), pp. 1–4 (2020). IEEE
126. El-Alfi, A., El-Gamal, A., El-Adly, R.: Real time arabic sign language to arabic text & sound translation system. Int. J. Eng **3**(5) (2014)
127. Salem, N., Alharbi, S., Khezendar, R., Alshami, H.: Real-time glove and android application for visual and audible arabic sign language translation. Procedia Computer Science **163**, 450–459 (2019)
128. Abraham, E., Nayak, A., Iqbal, A.: Real-time translation of indian sign language using lstm. In: 2019 Global Conference for Advancement in Technology (GCAT), pp. 1–5 (2019). IEEE
129. Escudeiro, N., Escudeiro, P., Soares, F., Litos, O., Norberto, M., Lopes, J.: Recognition of hand configuration: A critical factor in automatic sign language translation. In: 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–5 (2017). IEEE
130. Quach, L.-D., Duong-Trung, N., Vu, A.-V., Nguyen, C.-N.: Recommending the workflow of vietnamese sign language translation via a comparison of several classification algorithms. In: International Conference of the Pacific Association for Computational Linguistics, pp. 134–141 (2019). Springer
131. Xiaomei, Z., Shiquan, D., Hui, W.: Research on chinese-american sign language translation. In: 2011 14th IEEE International Conference on Computational Science and Engineering, pp. 555–558 (2011). IEEE
132. Liqing, G., Wenwen, L., Yong, S., Yanyan, W., Guoming, L.: Research on portable sign language translation system based on embedded system. In: 2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE), pp. 636–639 (2018). IEEE
133. Dajie, X., Shuning, K., Songlin, L.: Research on the translation of gloves based on embedded sign language. Digital Technology and Application (2017)
134. Singh, D.K., Kumar, A., Ansari, M.A.: Robust modelling of static hand gestures using deep convolutional network for sign language translation. In: 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 487–492 (2021). IEEE
135. Narashiman, D., Vidhya, S., Mala, D.T.: Tamil noun to sign language-a machine translation approach. In: Proceeding of 11th Tamil Internet Conference, pp. 175–179 (2012)
136. Alam, M., Tanvir, M., Saha, D.K., Das, S.K., et al.: Two dimensional convolutional neural network approach for real-time bangla sign language characters recognition and translation. SN Computer Science **2**(5), 1–13 (2021)
137. Zou, X., Chai, Y., Ma, H., Jiang, Q., Zhang, W., Ma, X., Wang, X., Lian, H., Huang, X., Ji, J., et al.: Ultrahigh sensitive wearable pressure sensors based on reduced graphene oxide/polypyrrole foam for sign language translation. Advanced Materials Technologies **6**(7), 2001188 (2021)
138. Sonare, B., Padgal, A., Gaikwad, Y., Patil, A.: Video-based sign language translation system using machine learning. In: 2021 2nd International Conference for Emerging Technology (INCET), pp. 1–4 (2021). IEEE
139. Madhuri, Y., Anitha, G., Anburajan, M.: Vision-based sign language translation device. In: 2013 International Conference on Information Communication and Embedded Systems (ICICES), pp. 565–568 (2013). IEEE
140. Lee, S., Jo, D., Kim, K.-B., Jang, J., Park, W.: Wearable sign language translation system using strain sensors. Sens. Actuators, A **331**, 113010 (2021)
141. Kim, T., Kim, S.: Sign language translation system using latent feature values of sign language images. In: 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 228–233 (2016). IEEE

142. Domingo, A., Akmeliawati, R., Chow, K.Y.: Pattern matching for automatic sign language translation system using labview. In: 2007 International Conference on Intelligent and Advanced Systems, pp. 660–665 (2007). IEEE
143. Yetkin, O., Calderon, K., Krishna Moorthy, P., Nguyen, T.T., Tran, J., Terry, T., Vigil, A., Alsup, A., Tekleab, A., Sancillo, D., *et al.*: A lightweight wearable american sign language translation device. In: Frontiers in Biomedical Devices, vol. 84815, pp. 001–04007 (2022). American Society of Mechanical Engineers
144. Kuriakose, Y.V., Jangid, M.: Translation of american sign language to text: Using yolov3 with background subtraction and edge detection. In: Smart Systems: Innovations in Computing, pp. 21–30. Springer, ??? (2022)
145. Shokoori, A.F., Shinwari, M., Popal, J.A., Meena, J.: Sign language recognition and translation into pashto language alphabets. In: 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1401–1405 (2022). IEEE
146. Bismoy, M.I., Shahrear, F., Mitra, A., Bikash, D., Afrin, F., Roy, S., Arif, H.: Image translation of bangla and english sign language to written language using convolutional neural network. In: 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), pp. 1–6 (2022). IEEE
147. Rajarajeswari, S., Renji, N.M., Kumari, P., Keshavamurthy, M., Kruthika, K.: Real-time translation of indian sign language to assist the hearing and speech impaired. In: Innovations in Computational Intelligence and Computer Vision, pp. 303–322. Springer, ??? (2022)
148. Dabhade, T., Ghawate, S., Diwane, A., Andrade, C., Chavan, P.: Sign language translation using cnn survey
149. Abougarair, A., Arebi, W.: Smart glove for sign language translation. Int Rob Auto J **8**(3), 109–117 (2022)
150. Wu, R., Seo, S., Ma, L., Bae, J., Kim, T.: Full-fiber auxetic-interlaced yarn sensor for sign-language translation glove assisted by artificial neural network. Nano-Micro Letters **14**(1), 1–14 (2022)
151. Klomsae, A., Auephanwiriyakul, S., Theera-Umpon, N.: A novel string grammar unsupervised possibilistic c-medians algorithm for sign language translation systems. Symmetry **9**(12), 321 (2017)
152. Zhou, Z., Neo, Y., Lui, K.-S., Tam, V.W., Lam, E.Y., Wong, N.: A portable hong kong sign language translation platform with deep learning and jetson nano. In: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, pp. 1–4 (2020)
153. Kau, L.-J., Su, W.-L., Yu, P.-J., Wei, S.-J.: A real-time portable sign language translation system. In: 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 1–4 (2015). IEEE
154. Park, H., Lee, J.-S., Ko, J.: Achieving real-time sign language translation using a smartphone's true depth images. In: 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), pp. 622–625 (2020). IEEE
155. Eqab, A., Shanableh, T.: Android mobile app for real-time bilateral arabic sign language translation using leap motion controller. In: 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA), pp. 1–5 (2017). IEEE
156. Tumsri, J., Kimpan, W.: Applied finite automata and quadtree technique for thai sign language translation. In: International MultiConference of Engineers and Computer Scientists, pp. 351–365 (2017). Springer
157. Kanvinde, A., Revadekar, A., Tamse, M., Kalbande, D.R., Bakereywala, N.: Bidirectional sign language translation. In: 2021 International Conference on Communication Information and Computing Technology (ICCICT), pp. 1–5 (2021). IEEE
158. Kaur, P., Ganguly, P., Verma, S., Bansal, N.: Bridging the communication gap: with real time sign language translation. In: 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), pp. 485–490 (2018). IEEE
159. Park, C.-I., Sohn, C.-B.: Data augmentation for human keypoint estimation deep learning based sign language translation. Electronics **9**(8), 1257 (2020)
160. Salim, B.W., Zeebaree, S.R.: Design & analyses of a novel real time kurdish sign language for kurdish text and sound translation system. In: 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S &T), pp. 348–352 (2020). IEEE
161. Lee, J., Heo, S., Baek, D., Park, E., Lim, H., Ahn, H.: Design and implementation of sign language translation program using motion recognition. International Journal of Hybrid Information Technology **12**(2), 47–54 (2019)
162. Hazari, S.S., Alam, L., Al Goni, N., *et al.*: Designing a sign language translation system using kinect motion sensor device. In: 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 344–349 (2017). IEEE
163. Putra, Z.P., Anasanti, M.D., Priambodo, B.: Designing translation tool: Between sign language to spoken text on kinect time series data using dynamic time warping. Sinergi (2018)
164. Pezzuoli, F., Tafaro, D., Pane, M., Corona, D., Corradini, M.L.: Development of a new sign language translation system for people with autism spectrum disorder. Advances in Neurodevelopmental Disorders **4**(4), 439–446 (2020)
165. Pezzuoli, F., Corona, D., Corradini, M.L., Cristofaro, A.: Development of a wearable device for sign language translation, 115–126 (2019)
166. Ab Majid, N.K., Norddin, N., Jaffar, K., Jaafar, R., Abd Halim, A.A., Ahmad, E.Z., dan Elektronik, F.T.K.E.: Development of a wearable glove for a sign language translation. Proceedings of Mechanical Engineering Research Day 2020, 263–265 (2020)
167. Neo, K.C., Ibrahim, H.: Development of sign signal translation system based on altera's fpga de2 board. International Journal of Human Computer Interaction (IJHCI) **2**(3), 101 (2011)
168. Park, H., Lee, Y., Ko, J.: Enabling real-time sign language translation on mobile platforms with on-board depth cameras. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **5**(2), 1–30 (2021)
169. Reis, L.S., de Araújo, T.M.U., Aguiar, Y.P.C., Lima, M.A.C.B.: Evaluating machine translation systems for brazilian sign language in the treatment of critical grammatical aspects. In: Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems, pp. 1–6 (2020)
170. Madushanka, A., Senevirathne, R., Wijesekara, L., Arunatilake, S., Sandaruwan, K.: Framework for sinhala sign language recognition and translation using a wearable armband. In: 2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer), pp. 49–57 (2016). IEEE
171. Estrada Jiménez, L.A., Benalcázar, M.E., Sotomayor, N.: Gesture recognition and machine learning applied to sign language translation. In: VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016, pp. 233–236 (2017). Springer
172. Verma, H.V., Aggarwal, E., Chandra, S.: Gesture recognition using kinect for sign language translation. In: 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), pp. 96–100 (2013). IEEE
173. Fu, Q., Shen, J., Zhang, X., Wu, Z., Zhou, M.: Gesture recognition with kinect for automated sign language translation. J. Beijing Normal Univ.(Nat. Sci.) 49(6), 586–587 (2013)
174. Nagpal, A., Singha, K., Gouri, R., Noor, A., Bagwari, A.: Hand sign translation to audio message and text message: A device.

In: 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 243–245 (2020). IEEE
175. Osman, M.N., Sedek, K.A., Zain, N.Z.M., Karim, M.A.N.A., Maghribi, M.: Hearing assistive technology: Sign language translation application for hearing-impaired communication, 1–11 (2020)
176. Jose, M.J., Priyadharshni, V., Anand, M.S., Kumaresan, A., Mo-hanKumar, N.: Indian sign language (isl) translation system for sign language learning. International Journal of Innovative Research and Development **2**(5), 358–367 (2013)
177. Wilson, B.J., Anspach, G.: Neural networks for sign language translation. In: Applications of Artificial Neural Networks IV, vol. 1965, pp. 589–599 (1993). SPIE
178. Raziq, N., Latif, S.: Pakistan sign language recognition and translation system using leap motion device. In: International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, pp. 895–902 (2016). Springer
179. Akmeliawati, R., Ooi, M.P.-L., Kuang, Y.C.: Real-time malaysian sign language translation using colour segmentation and neural network. In: 2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007, pp. 1–6 (2007). IEEE
180. Pansare, J., Rampurkar, K.S., Mahamane, P.L., Baravkar, R.J., Lanjewar, S.V.: Real-time static devnagri sign language translation using histogram. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) **2**(4), 1455–1459 (2011)
181. Praveena, S., Jayasri, C.: Recognition and translation of indian sign language for deaf and dumb people. International Journal Of Information And Computing Science **6** (2019)
182. He, S.: Research of a sign language translation system based on deep learning. In: 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), pp. 392–396 (2019). IEEE
183. Elsayed, E.K., Fathy, D.R.: Sign language semantic translation system using ontology and deep learning. International Journal of Advanced Computer Science and Applications **11** (2020)
184. Sharma, A., Panda, S., Verma, S.: Sign language to speech translation. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–8 (2020). IEEE
185. Harini, R., Janani, R., Keerthana, S., Madhubala, S., Venkatasubramanian, S.: Sign language translation. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 883–886 (2020). IEEE
186. Fernando, P., Wimalaratne, P.: Sign language translation approach to sinhalese language. GSTF Journal on Computing (JoC) **5**(1), 1–9 (2016)
187. Abe, M., Sakou, H., Sagawa, H.: Sign language translation based on syntactic and semantic analysis. Systems and computers in Japan **25**(6), 91–103 (1994)
188. Khan, M., Siddiqui, N., *et al.*: Sign language translation in urdu/hindi through microsoft kinect. In: IOP Conference Series: Materials Science and Engineering, vol. 899, p. 012016 (2020). IOP Publishing
189. Antony, R., Paul, S., Alex, S., et al.: Sign language translation system. International Journal of Scientific Research & Engineering Trends **6** (2020)
190. Shi, G., Li, Z., Tu, K., Jia, S., Cui, Q., Jin, Y.: Sign language translation system based on micro-inertial measurement units and zigbee network. Trans. Inst. Meas. Control. **35**(7), 901–909 (2013)
191. Ohki, M., Sagawa, H., Hataoka, N., Fujisawa, H.: Sign language translation system using pattern recognition and synthesis. Hitachi review **44**(4), 251–254 (1995)
192. Wu, C.-H., Chiu, Y.-H., Cheng, K.-W.: Sign language translation using an error tolerant retrieval algorithm. In: Seventh International Conference on Spoken Language Processing (2002)
193. Abiyev, R.H., Arslan, M., Idoko, J.B.: Sign language translation using deep convolutional neural networks. KSII Transactions on Internet and Information Systems (TIIS) **14**(2), 631–653 (2020)
194. Khan, S.A., Ansari, Z.A., Singh, R., Rawat, M.S., Khan, F.Z., Yadav, S.K.: Sign translation via natural language processing. population **4**, 5
195. Vachirapipop, M., Soymat, S., Tiraronnakul, W., Hnoohom, N.: Sign translation with myo armbands. In: 2017 21st International Computer Science and Engineering Conference (ICSEC), pp. 1–5 (2017). IEEE
196. Sapkota, B., Gurung, M.K., Mali, P., Gupta, R.: Smart glove for sign language translation using arduino. In: 1st KEC Conference Proceedings, vol. 1, pp. 5–11 (2018)
197. Chanda, P., Auephanwiriyakul, S., Theera-Umpon, N.: Thai sign language translation system using upright speed-up robust feature and c-means clustering. In: 2012 IEEE International Conference on Fuzzy Systems, pp. 1–6 (2012). IEEE
198. Chanda, P., Auephanwiriyakul, S., Theera-Umpon, N.: Thai sign language translation system using upright speed-up robust feature and dynamic time warping. In: 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), vol. 2, pp. 70–74 (2012). IEEE
199. Phitakwinai, S., Auephanwiriyakul, S., Theera-Umpon, N.: Thai sign language translation using fuzzy c-means and scale invariant feature transform. In: International Conference on Computational Science and Its Applications, pp. 1107–1119 (2008). Springer
200. Auephanwiriyakul, S., Phitakwinai, S., Suttapak, W., Chanda, P., Theera-Umpon, N.: Thai sign language translation using scale invariant feature transform and hidden markov models. Pattern Recogn. Lett. **34**(11), 1291–1298 (2013)
201. Tumsri, J., Kimpan, W.: Thai sign language translation using leap motion controller. In: Proceedings of the International Multiconference of Engineers and Computer Scientists, pp. 46–51 (2017)
202. Maharjan, P., Bhatta, T., Park, J.Y.: Thermal imprinted self-powered triboelectric flexible sensor for sign language translation. In: 2019 20th International Conference on Solid-State Sensors, Actuators and Microsystems & Eurosensors XXXIII (TRANSDUCERS & EUROSENSORS XXXIII), pp. 385–388 (2019). IEEE
203. Izzah, A., Suciati, N.: Translation of sign language using generic fourier descriptor and nearest neighbour. International Journal on Cybernetics and Informatics **3**(1), 31–41 (2014)
204. Mean Foong, O., Low, T.J., La, W.W.: V2s: Voice to sign language translation system for malaysian deaf people. In: International Visual Informatics Conference, pp. 868–876 (2009). Springer
205. Jenkins, J., Rashad, S.: Leapasl: A platform for design and implementation of real time algorithms for translation of american sign language using personal supervised machine learning models. Software Impacts **12**, 100302 (2022)
206. Natarajan, B., Rajalakshmi, E., Elakkiya, R., Kotecha, K., Abraham, A., Gabralla, L.A., Subramaniyaswamy, V.: Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation. IEEE Access **10**, 104358–104374 (2022)
207. Axyonov, A.A., Kagirov, I.A., Ryumin, D.A.: A method of multimodal machine sign language translation for natural human-computer interaction. Journal Scientific and Technical Of Information Technologies, Mechanics and Optics **139**(3), 585 (2022)
208. Chattopadhyay, M., Parulekar, M., Bhat, V., Raisinghani, B., Arya, S.: Sign language translation using a chrome extension

for google meet. In: 2022 IEEE Region 10 Symposium (TENSYMP), pp. 1–5 (2022). IEEE
209. Wilson, E.J., Anspach, G.: Applying neural network developments to sign language translation. In: Neural Networks for Signal Processing III-Proceedings of the 1993 IEEE-SP Workshop, pp. 301–310 (1993). IEEE
210. Wang, S., Guo, D., Zhou, W.-g., Zha, Z.-J., Wang, M.: Connectionist temporal fusion for sign language translation. In: Proceedings of the 26th ACM International Conference on Multimedia, pp. 1483–1491 (2018)
211. Guo, D., Zhou, W., Li, H., Wang, M.: Hierarchical lstm for sign language translation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
212. Guo, D., Tang, S., Wang, M.: Connectionist temporal modeling of video and language: a joint model for translation and sign labeling. In: IJCAI, pp. 751–757 (2019)
213. Guo, D., Wang, S., Tian, Q., Wang, M.: Dense temporal convolution network for sign language translation. In: IJCAI, pp. 744–750 (2019)
214. Elons, A.S., Ahmed, M., Shedid, H.: Facial expressions recognition for arabic sign language translation. In: 2014 9th International Conference on Computer Engineering & Systems (ICCES), pp. 330–335 (2014). IEEE
215. Wurm, S.: Finding the bones for the skeleton: A case of developing sign language translation practices. In: The Third Community Interpreting Research Seminar in Ireland (2011)
216. Fei, B., Jiwei, H., Xuemei, J., Ping, L.: Gesture recognition for sign language video stream translation. In: 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), pp. 1315–1319 (2020). IEEE
217. Boulares, M., Jemni, M.: Learning sign language machine translation based on elastic net regularization and latent semantic analysis. Artif. Intell. Rev. **46**(2), 145–166 (2016)
218. Werapan, W., Chotikakamthorn, N.: Improved dynamic gesture segmentation for thai sign language translation. In: Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP'04. 2004., vol. 2, pp. 1463–1466 (2004). IEEE
219. Wu, C., Pan, C., Jin, Y., Sun, S., Shi, G.: Improvement of chinese sign language translation system based on collaboration of arm and finger sensing nodes. In: 2016 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 474–478 (2016). IEEE
220. Tu, K., Pan, C., Zhang, J., Jin, Y., Wang, J., Shi, G.: Improvement of chinese sign language translation system based on multi-node micro inertial measurement unit. In: 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 1781–1786 (2015). IEEE
221. Pezzuoli, F., Corona, D., Corradini, M.L.: Improvements in a wearable device for sign language translation. In: International Conference on Applied Human Factors and Ergonomics, pp. 70–81 (2019). Springer
222. Sagawa, H., Sakiyama, T., Oohira, E., Sakou, H., Abe, M.: Prototype sign language translation system. In: Proceedings of IISF/ACM Japan International Symposium, pp. 152–153 (1994)
223. Song, P., Guo, D., Xin, H., Wang, M.: Parallel temporal encoder for sign language translation. In: 2019 IEEE International Conference on Image Processing (ICIP), pp. 1915–1919 (2019). IEEE
224. Feng, S., Yuan, T.: Sign language translation based on new continuous sign language dataset. In: 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), pp. 491–494 (2022). IEEE
225. Yin, Q., Tao, W., Liu, X., Hong, Y.: Neural sign language translation with sf-transformer. In: 2022 the 6th International Conference on Innovation in Artificial Intelligence (ICIAI), pp. 64–68 (2022)
226. Samonte, M.J.C., Guingab, C.J.M., Relayo, R.A., Sheng, M.J.C., Tamayo, J.R.D.: Using deep learning in sign language translation to text
227. Zhou, Z., Tam, V.W., Lam, E.Y.: A portable sign language collection and translation platform with smart watches using a blstm-based multi-feature framework. Micromachines **13**(2), 333 (2022)
228. Tang, S., Guo, D., Hong, R., Wang, M.: Graph-based multimodal sequential embedding for sign language translation. IEEE Transactions on Multimedia (2021)
229. Nunnari, F., España-Bonet, C., Avramidis, E.: A data augmentation approach for sign-language-to-text translation in-the-wild. In: 3rd Conference on Language, Data and Knowledge (LDK 2021) (2021). Schloss Dagstuhl-Leibniz-Zentrum für Informatik
230. Borgotallo, R., Marino, C., Piccolo, E., Prinetto, P., Tiotto, G., Rossini, M.: A multilanguage database for supporting sign language translation and synthesis. In: sign-langLREC2010, pp. 23–26 (2010). European Language Resources Association (ELRA)
231. Morrissey, S.: An asessment of appropriate sign language representation for machine translation in the heathcare domain. In: Sign Language Corpora: Linguistics Issues Workshop (2009). Citeseer
232. Halawani, S.M.: Arabic sign language translation system on mobile devices. IJCSNS International Journal of Computer Science and Network Security **8**(1), 251–256 (2008)
233. Kaczmarek, M., Filhol, M.: Assisting sign language translation: what interface given the lack of written form and the spatial grammar? In: Translating and the Computer (2019)
234. Baumgärtner, L., Jauss, S., Maucher, J., Zimmermann, G.: Automated sign language translation: The role of artificial intelligence now and in the future. In: CHIRA, pp. 170–177 (2020)
235. Morrissey, S., Somers, H., Smith, R., Gilchrist, S., Dandapat, S.: Building a sign language corpus for use in machine translation. Corpora and Sign Language Technologies, Representation and Processing of Sign Languages (2010)
236. Grif, M.G., Korolkova, O.O., Demyanenko, Y.A., Tsoy, E.B.: Computer sign language translation system for hearing impaired users. In: 2012 7th International Forum on Strategic Technology (IFOST), pp. 1–4 (2012). IEEE
237. Kaczmarek, M., Filhol, M.: Computer-assisted sign language translation: a study of translators' practice to specify cat software. Mach. Transl. **35**(3), 305–322 (2021)
238. Duarte, A.C.: Cross-modal neural sign language translation. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 1650–1654 (2019)
239. Kim, J., Hasimoto, A., Aoki, Y., Burger, A.: Design of a sign-language translation system between the japanese-korean by java-lifo language. In: Proceedings of IEEE. IEEE Region 10 Conference. TENCON 99.'Multimedia Technology for Asia-Pacific Information Infrastructure'(Cat. No. 99CH37030), vol. 1, pp. 423–426 (1999). IEEE
240. Ali, S.F., Mishra, G.S., Sahoo, A.K.: Domain bounded english to indian sign language translation model. International Journal of Computer Science and Informatics **3**(1), 41–45 (2013)
241. Ward, A., Escudeiro, N., Escudeiro, P.: Insights into the complexities of communication and automated sign language translation from the i-ace project. In: 2019 29th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE), pp. 1–5 (2019). IEEE
242. Hodorogea, V., et al.: Intersemiotics in contemporary advertising. from sign translation to meaning coherence. Professional Communication and Translation Studies (8), 45–55 (2015)
243. Kawano, S., Izumi, C., Kurokawa, T., Morimoto, K.: Japanese jsl translation and searching display conditions for expressing

easy-to-understand sign animation. In: International Conference on Computers for Handicapped Persons, pp. 667–674 (2006). Springer
244. Barberis, D., Garazzino, N., Prinetto, P., Tiotto, G., Savino, A., Shoaib, U., Ahmad, N.: Language resources for computer assisted translation from italian to italian sign language of deaf people. In: Proceedings of Accessibility Reaching Everywhere AEGIS Workshop and International Conference, pp. 96–104 (2011)
245. Kau, L.-J., Zhuo, B.-X.: Live demo: A real-time portable sign language translation system. In: 2016 IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 134–134 (2016). IEEE
246. Boulares, M., Jemni, M.: Mobile sign language translation system for deaf community. In: Proceedings of the International Cross-disciplinary Conference on Web Accessibility, pp. 1–4 (2012)
247. Wolfe, R., Efthimiou, E., Glauert, J., Hanke, T., McDonald, J., Schnepp, J.: recent advances in sign language translation and avatar technology. Univ. Access Inf. Soc. **15**(4), 485–486 (2016)
248. Liu, Z., Zhang, X., Kato, J.: Research on chinese-japanese sign language translation system. In: 2010 Fifth International Conference on Frontier of Computer Science and Technology, pp. 640–645 (2010). IEEE
249. Parton, B.S.: Sign language recognition and translation: A multidisciplined approach from the field of artificial intelligence. J. Deaf Stud. Deaf Educ. **11**(1), 94–101 (2006)
250. Wolfe, R.: Sign language translation and avatar technology. Mach. Transl. **35**(3), 301–304 (2021)
251. Grover, Y., Aggarwal, R., Sharma, D., Gupta, P.K.: Sign language translation systems for hearing/speech impaired people: a review. In: 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), pp. 10–14 (2021). IEEE
252. Camgöz, N.C., Varol, G., Albanie, S., Fox, N., Bowden, R., Zisserman, A., Cormier, K.: Slrtp 2020: The sign language recognition, translation & production workshop. In: European Conference on Computer Vision, pp. 179–185 (2020). Springer
253. Nguyen, T.B.D., Phung, T.-N.: Some issues on syntax transformation in vietnamese sign language translation. INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND NETWORK SECURITY **17**(5), 292–297 (2017)
254. Van Zijl, L., Olivrin, G.: South african sign language assistive translation. In: Proceedings of IASTED International Conference on Assistive Technologies, Page [no Page Numbers], Baltimore, MD (2008). Citeseer
255. Van Zijl, L., Barker, D.: South african sign language machine translation system. In: Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa, pp. 49–52 (2003)
256. Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., Abbott, S.: The development and evaluation of a speech-to-sign translation system to assist transactions. International Journal of Human-Computer Interaction **16**(2), 141–161 (2003)
257. Murtagh, I., Nogales, V.U., Blat, J.: Sign language machine translation and the sign language lexicon: A linguistically informed approach. In: Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track), pp. 240–251 (2022)
258. Jang, J.Y., Park, H.-M., Shin, S., Shin, S., Yoon, B., Gweon, G.: Automatic gloss-level data augmentation for sign language translation. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 6808–6813 (2022)
259. Shterionov, D., De Sisto, M., Vandeghinste, V., Brady, A., De Coster, M., Leeson, L., Blat, J., Picron, F., Scipioni, M., Parikh, A., *et al.*: Sign language translation: Ongoing development, challenges and innovations in the signon project. In: 23rd Annual Conference of the European Association for Machine Translation, pp. 323–324 (2022)
260. Huerta-Enochian, M., Lee, D.H., Myung, H.J., Byun, K.S., Lee, J.W.: Kosign sign language translation project: Introducing the niasl2021 dataset. In: Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives, pp. 59–66 (2022)
261. Efthimiou, E., Fotinea, S.-E., Hanke, T., McDonald, J.C., Shterionov, D., Wolfe, R.: Proceedings of the 7th international workshop on sign language translation and avatar technology: The junction of the visual and the textual: Challenges and perspectives. In: Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives (2022)
262. Bertin-Lemée, É., Braffort, A., Challant, C., Danet, C., Dauriac, B., Filhol, M., Martinod, E., Segouat, J.: Rosetta-lsf: an aligned corpus of french sign language and french for text-to-sign translation. In: 13th Conference on Language Resources and Evaluation (LREC 2022) (2022)
263. Kahlon, N.K., Singh, W.: Machine translation from text to sign language: a systematic review. Universal Access in the Information Society, 1–35 (2021)
264. Fang, B., Co, J., Zhang, M.: Deepasl: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation. In: Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems, pp. 1–13 (2017)
265. Guo, D., Zhou, W., Li, A., Li, H., Wang, M.: Hierarchical recurrent deep fusion using adaptive clip summarization for sign language translation. IEEE Trans. Image Process. **29**, 1575–1590 (2019). https://doi.org/10.1109/TIP.2019.2941267
266. Xu, W., Ying, J., Yang, H., Liu, J., Hu, X.: Residual spatial graph convolution and temporal sequence attention network for sign language translation. Multimedia Tools and Applications, 1–25 (2022)
267. Gu, Y., Zheng, C., Todoh, M., Zha, F.: American sign language translation using wearable inertial and electromyography sensors for tracking hand movements and facial expressions. Frontiers in Neuroscience **16** (2022)
268. Zhang, Q., Jing, J., Wang, D., Zhao, R.: Wearsign: Pushing the limit of sign language translation using inertial and emg wearables. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **6**(1), 1–27 (2022)
269. Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J.H., Ney, H.: Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus. In: LREC, vol. 9, pp. 3785–3789 (2012)
270. Othman, A., Jemni, M.: English-asl gloss parallel corpus 2012: Aslg-pc12. In: 5th Workshop on the Representation and Processing of Sign Languages: Interactions Between Corpus and Lexicon LREC (2012)
271. Hilzensauer, M., Krammer, K.: A multilingual dictionary for sign languages:"spreadthesign". ICERI2015 Proceedings, 7826–7834 (2015)
272. Huang, J., Zhou, W., Zhang, Q., Li, H., Li, W.: Video-based sign language recognition without temporal segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
273. Esplà-Gomis, M., Forcada, M., Ramírez-Sánchez, G., Hoang, H.T.: Paracrawl: Web-scale parallel corpora for the languages of the eu. In: MTSummit (2019)
274. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). Ieee

275. Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., et al.: The kinetics human action video dataset. arXiv preprint arXiv:1705.06950 (2017)
276. Li, D., Rodriguez, C., Yu, X., Li, H.: Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1459–1469 (2020)
277. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
278. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. J. Comput. Vision **1**(4), 321–331 (1988)
279. Luong, T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1412–1421. Association for Computational Linguistics, Lisbon, Portugal (2015). https://doi.org/10.18653/v1/D15-1166
280. Hanke, T., Schulder, M., Konrad, R., Jahn, E.: 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, pp. 75–82. European Language Resources Association (ELRA), Marseille, France (2020). https://aclanthology.org/2020.signlang-1.12
281. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A.M.: Transformers: State-of-the-Art Natural Language Processing (2020). https://www.aclweb.org/anthology/2020.emnlp-demos.6
282. Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., Zettlemoyer, L.: Multilingual denoising pre-training for neural machine translation. Transactions of the Association for Computational Linguistics **8**, 726–742 (2020)
283. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607 (2020). PMLR
284. Callison-Burch, C., Osborne, M., Koehn, P.: Re-evaluating the role of BLEU in machine translation research. In: 11th Conference of the European Chapter of the Association for Computational Linguistics (2006)
285. De Meulder, M.: Is "good enough" good enough? ethical and responsible development of sign language technologies. In: Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL), pp. 12–22. Association for Machine Translation in the Americas, Virtual (2021). https://aclanthology.org/2021.mtsummit-at4ssl.2
286. Vanmassenhove, E., Shterionov, D., Way, A.: Lost in translation: Loss and decay of linguistic richness in machine translation. In: Proceedings of Machine Translation Summit XVII Volume 1: Research Track, pp. 222–232. European Association for Machine Translation, Dublin, Ireland (2019). https://www.aclweb.org/anthology/W19-6622
287. Vanmassenhove, E., Shterionov, D., Gwilliam, M.: Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL, pp. 2203–2213 (2021). Association for Computational Linguistics. https://aclanthology.org/2021.eacl-main.188/
288. De Coster, M., Van Herreweghe, M., Dambre, J.: Isolated sign recognition from rgb video using pose flow and self-attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3441–3450 (2021). https://doi.org/10.1109/CVPRW53098.2021.00383
289. Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Pires, B., Guo, Z., Azar, M., *et al.*: Bootstrap your own latent: A new approach to self-supervised learning. In: Neural Information Processing Systems (2020)
290. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the International Conference on Computer Vision (ICCV) (2021)
291. Baevski, A., Zhou, Y., Mohamed, A., Auli, M.: wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in Neural Information Processing Systems **33** (2020)
292. Barrault, L., Bojar, O., Costa-Jussa, M.R., Federmann, C., Fishel, M., Graham, Y.: Findings of the 2019 conference on machine translation (wmt19). (2019). Association for Computational Linguistics (ACL)